TM-4743/012/01

AD 744670

REPORT OF THE CACTOS PROJECT:

A PRELIMINARY INVESTIGATION OF COMPUTATION AND COMMUNICATION TRADE-OFFS IN MILITARY COMMAND AND CONTROL SYSTEMS

N. E. WILLMORTH

1 APRIL 1972

149

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) System Development Corporation 2500 Colorado Avenue Santa Monica, California 90406	2a. REPORT SECURITY CLASSIFICATION 2b. GROUP

3 REPORT TITLE

Report of the CACTOS Project: A Preliminary Investigation of Computation and Communication Trade-Offs in Military Command and Control Systems

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical

5 AUTHOR(S) (First name, middle initial, last name)

Norman E. Willmorth

6 REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1 April 1972	ix, 142	29

8a. CONTRACT OR GRANT NO. DAHC15-67-C-0149 b. PROJECT NO. 2D30 c. d	9a. ORIGINATOR'S REPORT NUMBER(S) TM-4743/012/01 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11 SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	ARPA/IPT

13 ABSTRACT

This paper ~~reports~~ the progress of the Computation and Communication Trade-Off Study (CACTOS), conducted for the Advanced Research Projects Agency, for the purpose of determining the cost effectivity of new computer hardware, software, and communication channels for future Department of Defense requirements. Efforts to conceptualize DoD information processing needs and to develop analytic models and programs are reported. Technological alternatives are examined. A network analysis model is described.

This document is a reissue and replacement of TM-4743/012/00, which was a draft.

DD FORM 1 NOV 65 1473

ia

Unclassified

Security Classification

Unclassified

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
AUTODIN						
CACTOS Model						
Computer Network Analysis						
DESIGNET						
Dept. of Defense Computation/Communication Requirements 1975-1980						

Unclassified

Security Classification

TM-4743/012/01

REPORT OF THE CACTOS PROJECT:

A PRELIMINARY INVESTIGATION OF COMPUTATION AND COMMUNICATION TRADE-OFFS IN MILITARY COMMAND AND CONTROL SYSTEMS

N. E. WILLMORTH

1 APRIL 1972

THIS REPORT WAS PRODUCED BY SDC IN PERFORMANCE OF CONTRACT NO DAHC15 67 C 0141 ARPA ORDER NO 1327 AMENDMENT NO 4, PROJECT NO 2D30, PROGRAM CODE NO 2P10

ABSTRACT

This paper reports the progress of the Computation and Communication Trade-Off Study (CACTOS), conducted for the Advanced Research Projects Agency, for the purpose of determining the cost effectivity of new computer hardware, software, and communication channels for future Department of Defense requirements. Efforts to conceptualize DoD information processing needs and to develop analytic models and programs are reported. Technological alternatives are examined. A network analysis model is described.

This document is a reissue and replacement of TM-4743/012/00, which was a draft.

TABLE OF CONTENTS

Page

FIGURES

FIGURES (Cont'd)

TABLES

1. INTRODUCTION

The goal of the Computation and Communication Trade-Off Study (CACTOS) is to determine future Department of Defense requirements for computation and communication networks on regional, functional, and categorical levels on the basis of operational needs, technological development, and cost effectiveness

The DoD, as the largest single user of computation and communication equipment in the world, has acquired computers of all sizes and varieties, many of which are no longer cost effective for their assigned tasks. This situation has evolved: first, because the data processing load has grown beyond the capacity of the older computers; second, because new data processing concepts have obsoleted many once very sophisticated systems; and, third, because rapidly developing technology has resulted in faster, cheaper equipment.

To deal with the new data processing loads, a concept of operation reflecting modern trends toward interactive computer utilities is being adopted--many small computers will be replaced by single, larger, faster processors; functions will be integrated; and many more applications will be automated. However, the task of determining the precise nature and configuration of computers that will perform the required functions and provide for future growth in a maximally cost-effective manner is enormous. To satisfy the demands for advanced computers, equally advanced communications are required. Until recently pressed by increasing digital data demands, communication technology has been slower to evolve. However, DoD has been an active instigator of advanced communication system development and a leader in developing communication satellites, microwave installations, low-frequency systems, and single-side-band radio. Unfortunately, most of this arsenal consists of voice-grade channels that are only marginally useful in view of the very high reliability and low noise requirements of digital communications at high data rates.

One of the most momentous developments of our times has been the marriage of the computer to communication lines. While the communication lines provide the means of collecting and distributing the huge amounts of information and performing the computing tasks demanded of modern computing machinery, the computer in turn provides the control capability necessary to direct traffic through the networks of communication channels. The computer can perform other communication tasks, such as buffering between lines and communication devices of differing speeds, coding schemes, and message formats and performing intricate calculations for the demodulation and encryption of signals. This is truly a symbiotic relationship, for neither the computing system nor the communication system could have reached its current level of power without the support of the other.

One of the first large communication system to use computers in this way was AUTODIN, the DoD's automatic digital information system, which uses complexes of computers for message buffering, message switching, and line control. Although AUTODIN now interfaces with satellite, microwave, and other high-speed lines, the operating speeds and capabilities of its internal lines are relatively slow and limited in relation to potentially available communications capabilities. Very-high-speed lines; more cost-effective computers; faster, better switches; and new concepts for digital data communications could enable large volumes of messages to be multiplexed onto high-speed digital lines and reconstituted and distributed through a sequence of computerized, store-and-forward communication centers. Although many actions (such as adding more and faster lines and improving terminal buffering of voluminous data) have been taken to keep AUTODIN abreast of the growing DoD communication needs, the time is approaching when AUTODIN, too, should be replaced.

DoD has also pioneered in the development of on-line systems for command and control. The technical success of these systems, combined with the tremendous appeal of the computer utility concept (delivering the power of a large, high-speed computer to anyone possessing compatible control and display devices),

suggests a solution to the problems inherent in many of the mixed computation and communication functions of DoD. Although a larger, more efficient central computer may perform computing tasks more economically, many trade-offs are possible in designing a computation and communication system. Any piece of computing equipment has advantages and disadvantages for performing different tasks, and different task mixes may favor different constellations of computers and communication gear. Decisions concerning computer and communication equipment replacements involve millions of dollars and have tremendous impact upon the military establishment, the government, and the public. Understanding the functional relations involved in trade-offs and having a rational system-replacement strategy is essential. In addition, studying the problems involved in formulating trade-offs and replacement strategies develops opportunities for advancing the state of the computing and communication arts and for creating methods and procedures for handling such problems in the future.

That DoD is aware of the problems attendant upon the marriage, or netting, of computers and digital data communications is evident in the recent creation of the post of Assistant to the Secretary of Defense (Telecommunications) with responsibilities that include establishing standards to ensure interface compatibility for command and control systems. The 1970 report of the Blue Ribbon Defense Panel also stressed the importance of computer procurement and application. Neither action fully expresses the essential interdependence of computation and communication but indicates recognition of the role of computers and digital data in communication. How netted computers relate to non-netted ones in covering the total computation task is not known. If trade-off studies should establish the valid interrelation of computers and communications in handling the total computation and data transmission load, further weight would be given to the integrated management of such systems. It is possible, however, that the technological explosion in digital data processing capabilities may be creating a mismatch between ever-increasing demands for data transmission and the capacity of existing and planned communication systems.

Technical questions are raised by this rapid integration of digital data processing and transmission capabilities. Are distributed data bases more efficient than centralized files? Is message switching more efficient than circuit switching? Under what conditions of use is one solution better than another? Whatever the answers to these questions, the growth of data processing is having a major impact upon existing communication facilities. The communication demands of the total complex of computing facilities must be assessed in view of future possibilities for computation and communication systems. Of concern to DoD engineers and managers is the relative cost-effectiveness of the options and trade-offs available to them and the actions they might take to prevent a runaway technology from obsoleting systems faster than modifications can maintain cost-effective operations.

Obviously, some means are needed to determine the critical factors, to avoid the most stringent effects of the obsolescence rate, to keep costs within reason, to identify the facilitating and damping factors, and to avoid the possible pitfalls and impasses of improper matches between interdependent sets of requirements and developments.

The cost of modifying a million bits (one megabit) through the central processing unit of a computer has been plotted over a 15-year period since the delivery of large computers in 1956 (Figure 1). The costs of computation decrease dramatically by an order of magnitude every six years. Development of digital-data-transmission technology is also occurring rapidly (Figure 2), but the economic and physical construction constraints imposed by large and complex systems of communication lines enforces a more conservative implementation process than is required by ever more compact computing equipment.

Equally great strides are being made in applying this potential power to command and control and other information systems throughout DoD (Figure 3).

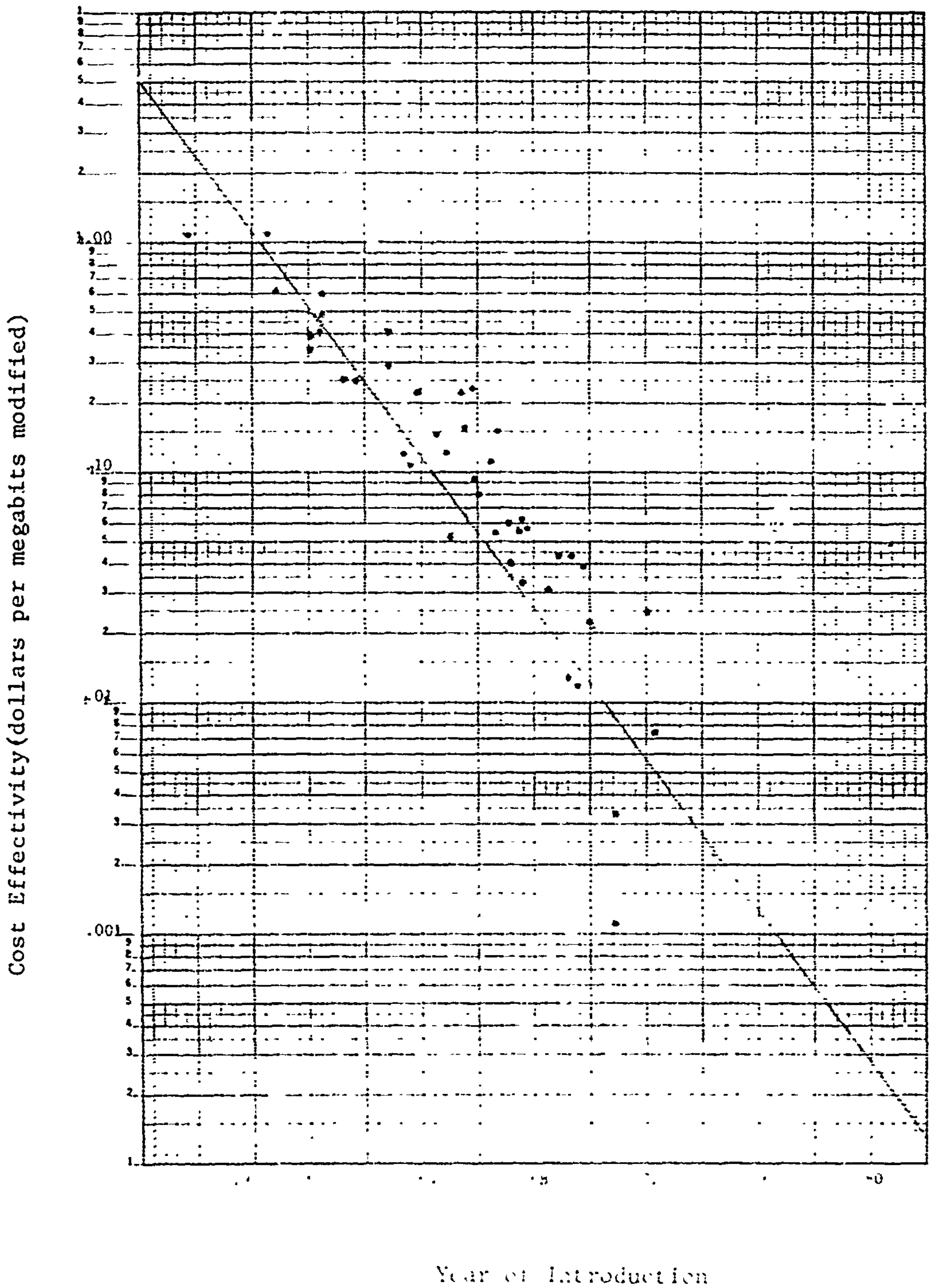


Figure 1. Computation Cost Effectivity: An Order of Magnitude Change Approximately every six Years

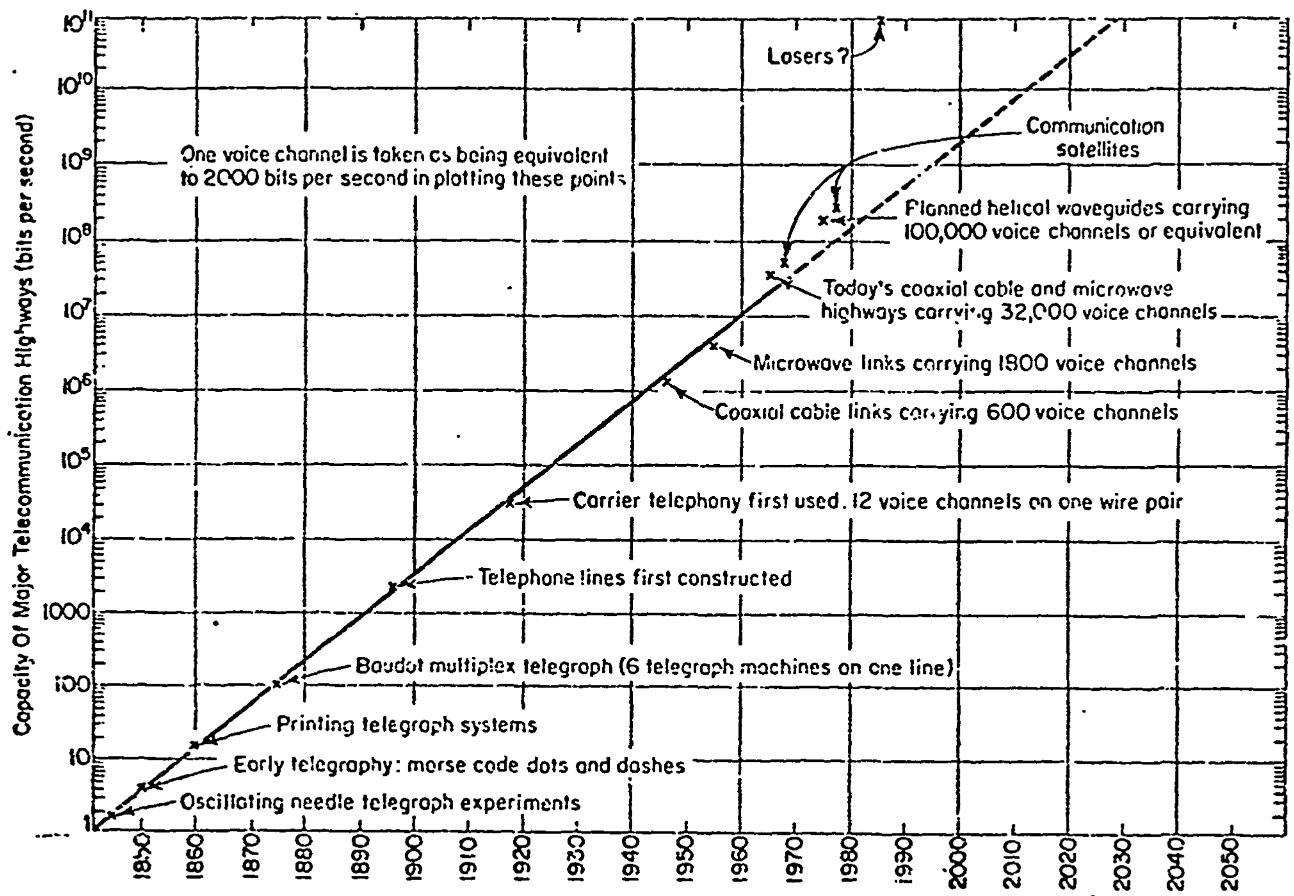


Figure 2. History of Communication Technology*

*From J. Martin, Telecommunications and the Computer. Prentice-Hall, Inc., 1969 p. 8.

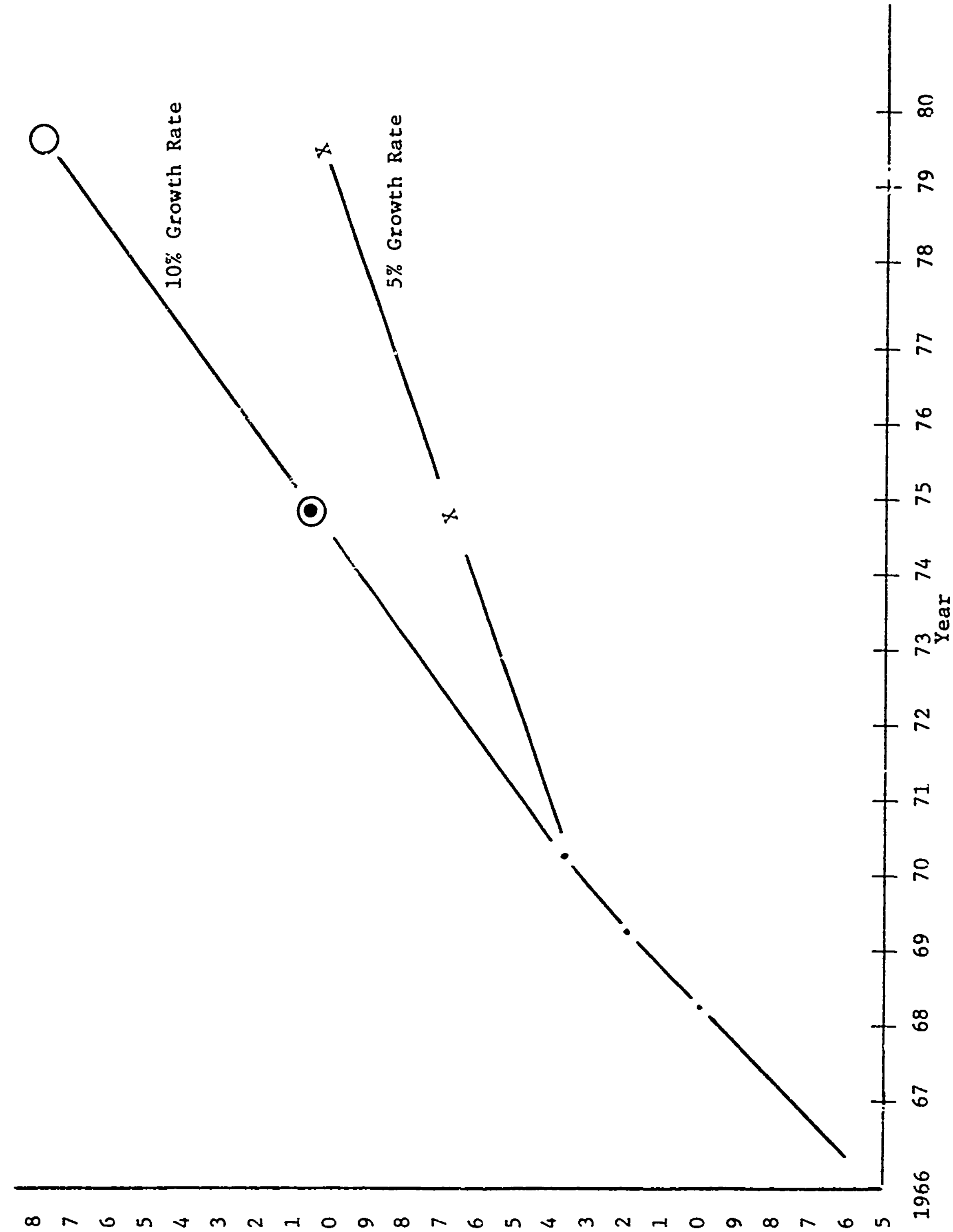


Figure 3. Projections for the Growth of Digital Data Processing at 5 and 10% Levels

This study proposes to provide, through an amalgam of recently developed techniques, some greater perspective upon the present explosive growth of computer employment and the implications of growth rates for the last several years, to project these into the future, and, then, to ask whether communications are being given adequate attention and whether the new systems will be economically feasible to procure. The study will attempt to treat the total problem of the interrelation of computation and communication and to seek a best path toward future systems through cost-effective trade-offs. In short, the project will take the systems approach toward determining what is needed (the "forcing factor" of DoD computation and communication requirements), whether a technology exists for providing it, and, if not, what actions should be taken to achieve it. An overview of the CACTOS approach is given in Figure 4.

The study is based upon these assumptions:

- DoD's requirements for computational capacity will continue to grow at a rate at least commensurate with current growth.

- Current computer acquisition plans, both for much-needed replacements for technologically obsolete machines and for new, advanced systems, imply a tremendous increase in digital data transmission needs, much of which will have to be handled by what now seems a conservative data transmission system.

- Although digital traffic is only a fraction of the total communications load, the digital traffic required to support an ever more complex, far-flung, and technologically sophisticated military establishment is growing at an even faster rate than analog communications.

- Computation and communication are coordinate activities whose influence upon one another is crucial, and they cannot be considered separately.

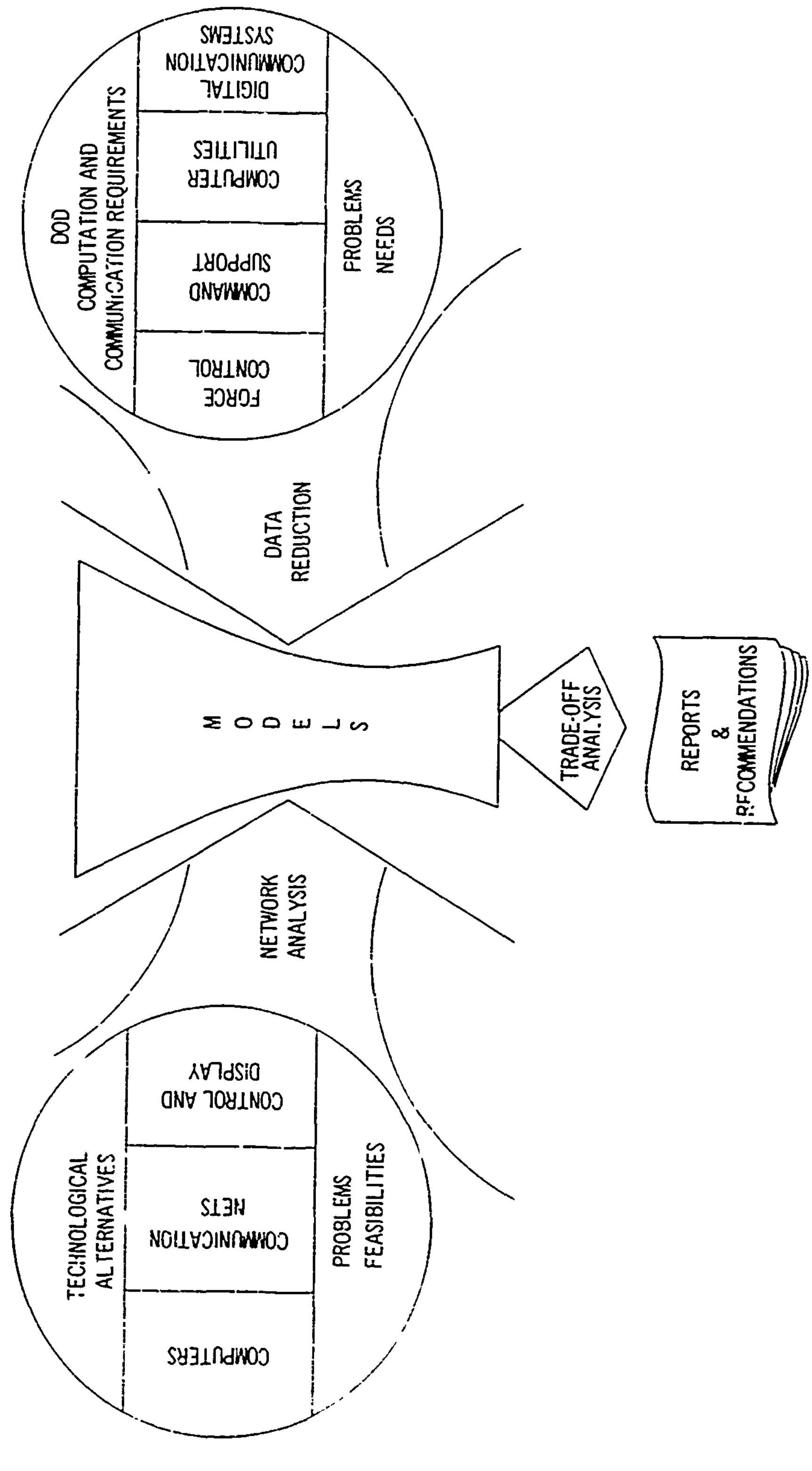


Figure 4. The CACTOS Approach

- A rational means must be found of keeping pace in a cost-effective manner with technological developments.

If these assumptions are valid, it may be concluded that a crisis exists or is building. In the civilian sector, the belief in the crisis is reflected in the very active development of digital data transmission systems, such as DATRAN, that are now challenging the established public utilities in providing digital data communication services. Not everyone believes that a crisis exists. One representative of Bell Telephone Laboratories feels that the normal technological growth of computational and communication equipment will be able to handle whatever burden is placed upon it (Hough, 1970). This may be true, but if the computation/communication mismatch does exist, it will have exceedingly serious consequences not only for the military establishment but for the nation as a whole. While actual breakdowns in operations may not occur, in consideration of the lead times required to determine requirements and procure a major system, Some answers to the computation/communication dilemma must be found. DoD has traditionally acted to control technological advance to its advantage by supporting those items of research and development that are crucial to support its needs, and it should continue to do so. Certainly, while technology holds promise of meeting almost any challenge, it will require vision to select the appropriate solution and to ensure that the solution is available when needed.

On the basis of these premises, CACTOS intends to:

- Develop a methodology for describing and analyzing computation and communication systems.
- Determine DoD information processing and transmission needs, especially as these apply to specific operational needs and functions.

- Investigate the cost-effectiveness of various technological trade-offs in alternative computation and communication network configurations and methods in view of evolving technological and economic factors.
- Develop optimal planning policies for the design of computation and communication networks and for the incorporation of the evolving technology into DoD systems.

In its first year, the project has:

- Surveyed technological and economic developments in the information processing sciences to detect areas of crucial impact upon the cost-effectiveness of future systems.
- Developed a prototype network analysis program.
- Formulated some preliminary computation and communication trade-offs.
- Validated the network analysis model using an existing system.
- Conducted some preliminary analyses of theoretical trade-offs in network behavior and computer throughput.

This report summarizes the progress that has been made on the project. Section 2 reports the results of efforts to conceptualize DoD information processing needs and to develop computation and communication network analysis models and programs. Section 3 summarizes an evaluation of the technological alternatives that might provide the basis for profitable trade-offs. Section 4 reports the results of the experimentation that has been done in validating the network analysis model and in preliminary investigations of trade-offs. Section 5 draws some preliminary conclusions and makes recommendations for future investigations.

2. TECHNICAL DEVELOPMENT

2.1 COMPUTATION AND COMMUNICATION REQUIREMENTS

DoD computation and communication requirements are dictated by four general categories of information processing activities. These are force control, command support, computer utilities, and communications.

The tasks of force-control information systems range from the monitoring and control of individual combat and support vehicles and squads, through tactical surveillance of battlefield situations, status of forces, and orders of battle, to the strategic planning of operations and force deployment. The information problems inherent in operational, tactical, and strategic planning and control systems are formidable. There are a large variety of sensors and other data collection devices whose detections must be screened and evaluated. Millions of decisions must be made that require the integration and display of a multitude of interrelated data. Additionally, the system must meet severe requirements for response time, security, accuracy, and continuity of operation.

The tasks of command-support information systems involve the collection, storage, reduction, and dissemination of data on weather, intelligence, logistics, personnel, and finance operations. While response-time requirements are less stringent than they are for force control (except in direct support of operations), the need for huge stores of both current and historical data, for the precise and accurate retrieval of information, and or the assessment of trends and the interrelation of millions of bits of information creates problems that are equally difficult.

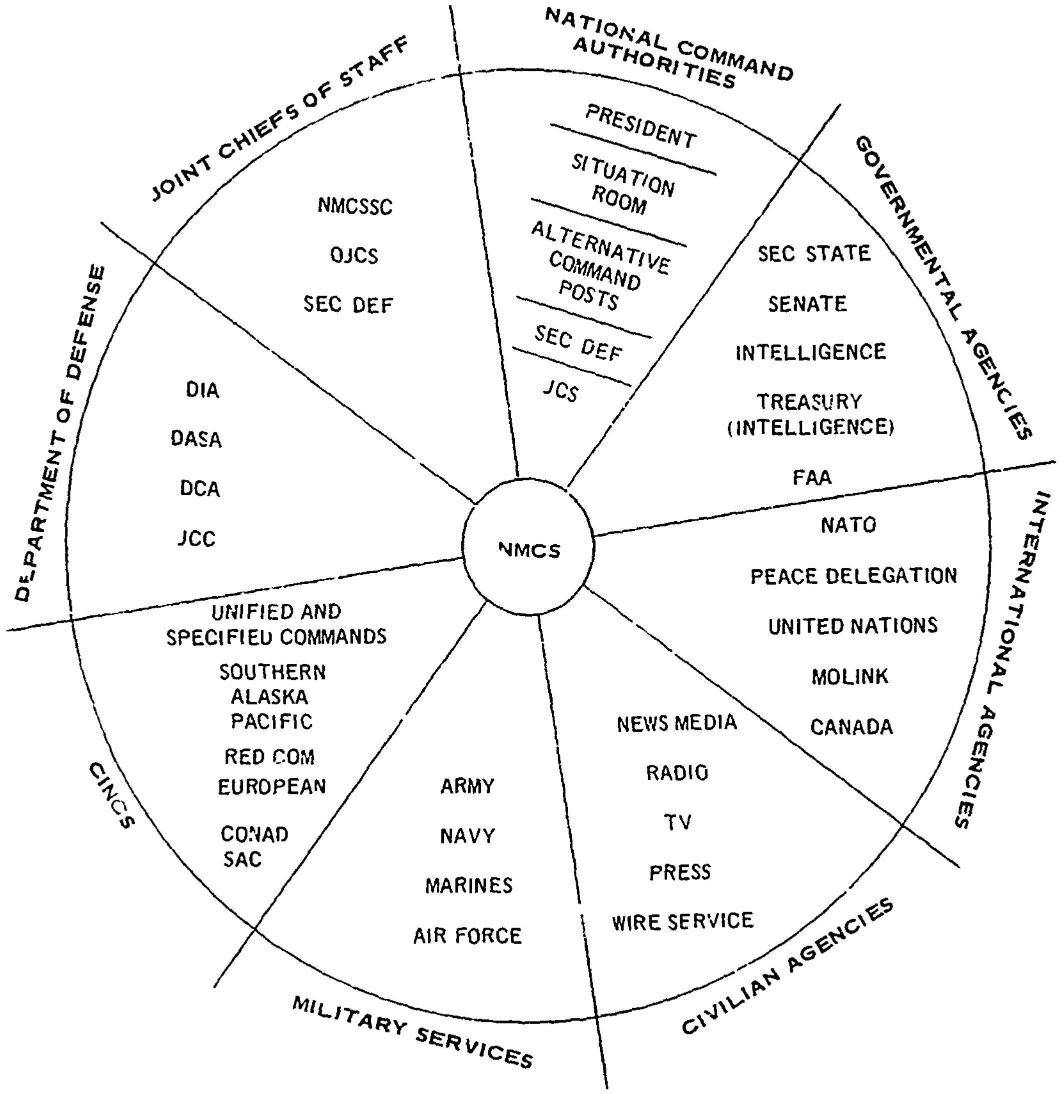
Most computing facilities handle a variety of tasks that are individually too small to justify a system dedicated to a single task. These tasks include scientific (mathematical and statistical) data processing, business applications, and information storage and retrieval. With the economies of scale inherent in high-capacity computation and communication equipment, private, public, and

military users are establishing larger and more highly interconnected computer utilities. These utilities have special requirements for versatility, economy, and operating efficiency that are not always met by the operating systems that control data processing and transmission devices. In the military, many base-level and area-management systems operate in the computer utility mode.

By far the greatest volume of DoD information to be processed is found in verbal and pictorial communication. Most "pure" communication systems are more analog than digital. However, in the civilian sector, commerical communication agencies are rapidly converting high-speed trunk lines to digital operation (i.e., pulse-code modulation (PCM) and time-division multiplexing (TDM)). Many military agencies are also interested in all-digital systems, as evidenced by the now defunct MALLARD system. An all-digital system has much to offer in the way of reduced error rates, improved security, circuit simplicity, and economy of operation in addition to avoiding the necessity of separate networks to handle analog and digital traffic. Processing requirements for communications-only systems are relatively low, but they are greater for storage, format conversion, and message handling in a store-and-forward network.

2.1.1 The Scope of DoD Systems

The scope of the computation and communication systems operated by the Department of Defense is illustrated in Figures 5 and 6. Many agencies and areas have one or more data processing systems that interface with the National Military Command System. At the top of the military hierarchy are the National Military Command Authorities--the President, the Secretary of Defense, and the Joint Chiefs of Staff--with several situation rooms and alternative command posts to be supported. There are then many agencies at the DoD/JCS level that are almost wholly engaged in information manipulation--gathering and evaluating information and planning. Each service has huge establishments to manage, and the joint and unified commands have large areas and many diverse units and operations to control. In addition, DoD systems interface with other governmental, non-military systems engaged in international relations, intelligence-data processing,



OVER 100 COMMAND AND CONTROL SYSTEMS

HUNDREDS OF SUPPORT COMPUTERS

OVER 3000 COMPUTERS TODAY

GROWING AT 10 – 15% PER YEAR

AUTODIN (CONUS) ALONE HAS 108 2400-BPS LINES
DEDICATED SYSTEMS HAVE THOUSANDS MORE

Figure 5. The World of DoD Computation and Communication Systems.

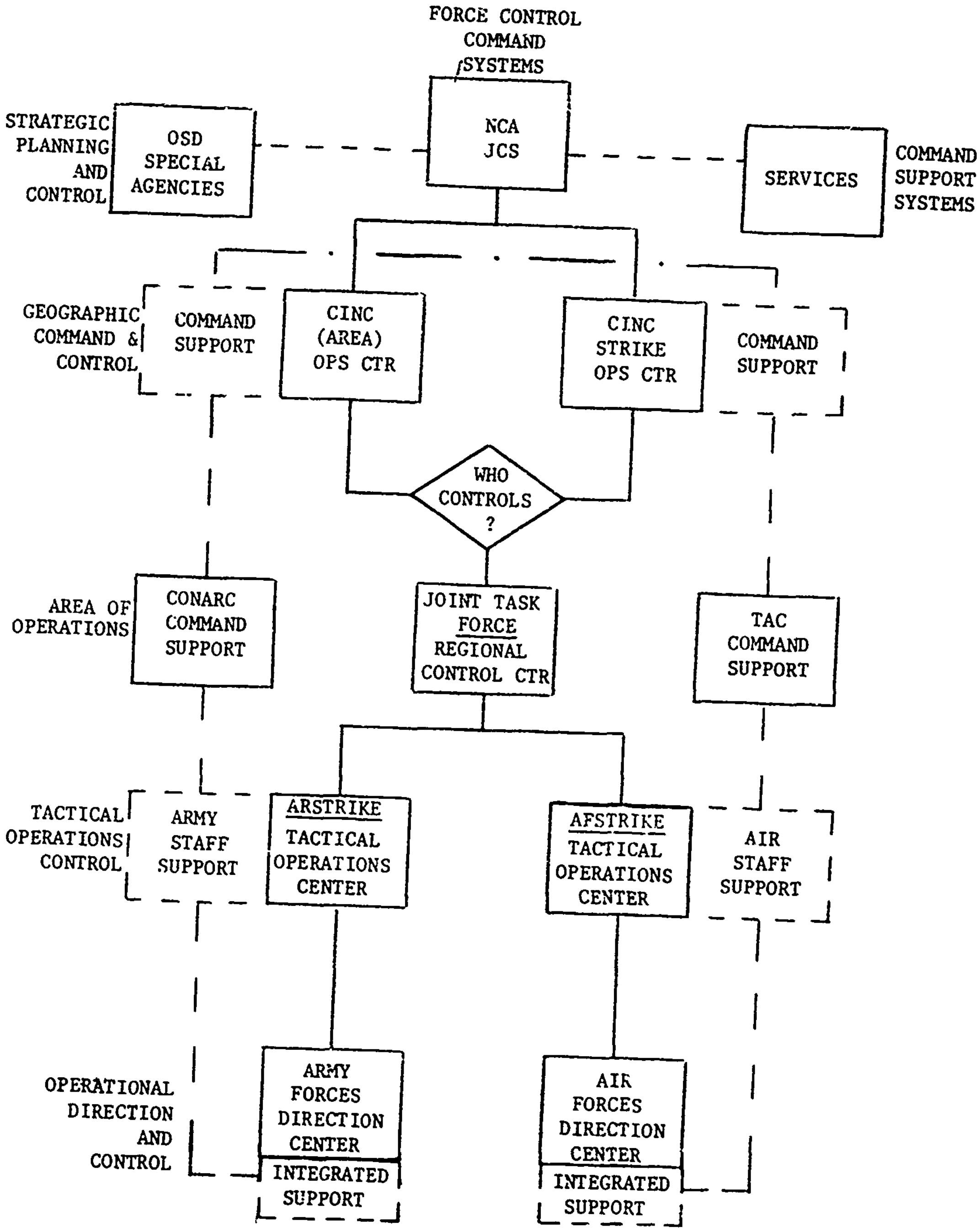


Figure 6. The Computation and Communication Systems Hierarchy

air traffic control, environmental control, communications, and other pursuits of joint interest. Communication interfaces exist with news media, the United Nations, other countries, and special delegations for peace and arms negotiations.

To sum up, the military establishment now has more than 100 command and control systems and thousands of independent computers. CONUS AUTODIN alone has 25 trunk lines in the 4800-bps range. Converted to AUTODIN worldwide are about 1300 teletype and high-speed terminals. AUTODIN serves about 300,000 users via 17,000 access lines interconnected by a network of approximately 7,000 trunks. There are many thousands of dedicated communication lines outside the integrated communication systems. Although DoD now owns or leases more than 3000 computers and is acquiring 10-15% more each year, most of these are in the small-to-medium, first- and second-generation classes. For improved cost effectiveness, these small and technologically obsolete computers should be replaced or consolidated with larger, technologically more advanced machines that take advantage of the economies of scale, specialization, and quality.

An idea of the depth of the computation and communication systems within DoD can be realized by following the flow of information through the system, beginning at the bottom with direct control of forces in accomplishing a mission (Figure 6). Detailed situational information and force-status information, in conjunction with specific mission objectives and a plan of attack, are used in making operational decisions. Summaries of these and the changes they create in the situation and force status are forwarded to tactical operations centers for integrated displays and further decision-making involving larger forces, areas, and missions. At the direct operational level, command-support operations are frequently an indistinguishable part of the total operation, but become staff support at tactical and higher levels.

At the top of the hierarchy, most information processing is done for analysis and planning; little of it is currently automated. Files of information on

historical trends, capabilities, economic and technical possibilities, and budgetary affairs are often available for interrogation. Staffs of specialists and assistants collect and evaluate information and prepare plans. The involvement with operational details is minimal except in emergency situations. Decisions are long range and broad, and the information that rises from the lower echelons is largely administrative detail requiring little in the way of command decisions. Information flowing down is largely concerned with plans and their interpretation. Command decisions, even in emergencies, normally involve implementing one of several alternative contingency plans. A decision thus made would set in motion the mobilization of forces (the creation of a Joint Task Force, for instance) and the logistical operations to support the mobilization.

Beneath the command/control authorities are the area commanders-in-chief. Much of the data processing at this level involves the administrative details of running an organization covering a large geographic area. Here, also, decisions are mostly concerned with plans, budgets, and schedules, but the potential for the relatively close control of vast forces must exist.

Although the distinction between the regional and specific mission may not always be clear, the nature of force-control decisions shifts in orientation from one to the other with the creation of a Joint Task Force. Strategies, plans, and schedules may still be important, but the goals are of much shorter range, and tactical force-control and support give rise to most of the information dealt with.

2.1.2 Command and Control Network Configurations

The required flow of information within a network usually determines the configuration of nodes and links. In a general communication system with minimal computing capability (i.e., that required for accounting and traffic control), the normal configuration is a clustering of lines from terminals around a local switching and concentration center in a star net with interconnected subnets of

switching centers and multiplexed trunk lines. In a computer utility, the normal configuration has a powerful computer in the middle of a star net arrangement of computers. If long distances are involved, terminals may be clustered around concentrators in a star net or multidrop configuration with one or more sets of trunk lines and concentration points carrying information to and from the computer. The concentrators may have switching capability for interterminal communication and also limited computing capability. Hence, there are numerous network alternatives and trade-offs to be evaluated.

Although there are few, if any, force-control systems whose data processing procedures have been automated throughout the command hierarchy, the general configuration of force-control systems is hierarchical, with several interconnected layers of computers integrating input data, displaying it to control-center personnel for command decisions, transmitting the decisions to the forces controlled, and sending situation summaries and decision referrals to high-command centers. Some communication is required between the nodes of a given level of a net, especially to pass control, to balance the load, or to compensate for the destruction or degradation of adjacent nodes. The principal flow of information, however, is up and down the chain of command. Figure 7 depicts the generalized force-control system network.

Command-support systems have operated in the past either as part of a local computer utility or as dedicated computers for largely local support operations. The present trend is toward large, centralized computers and toward distributed networks of computers to handle both operational and managerial data. Typical of the future trend are the proposed logistics systems of the various services. Figure 8 depicts such a distributed network. Within the figure are represented base-level operations, depot and area depots, local loop and regional networks, and headquarters operations as some indication of the potential complexity of the systems. Determining the proper distribution of transmission, computing, and storage capacity for such a system is a prime example of computation and communication trade-off analysis.

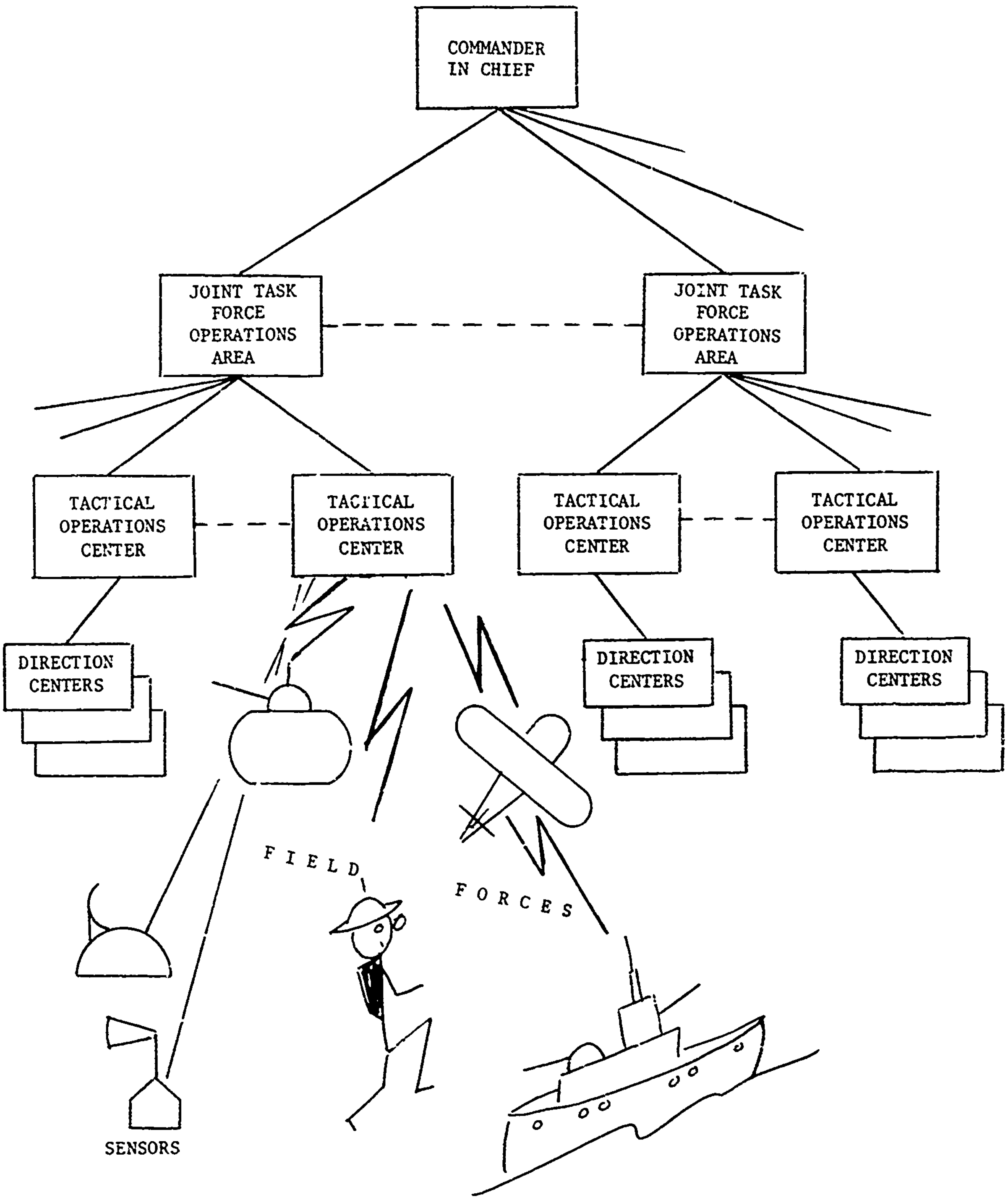


Figure 7. The Typical Hierarchical Configuration of a Force Control Network

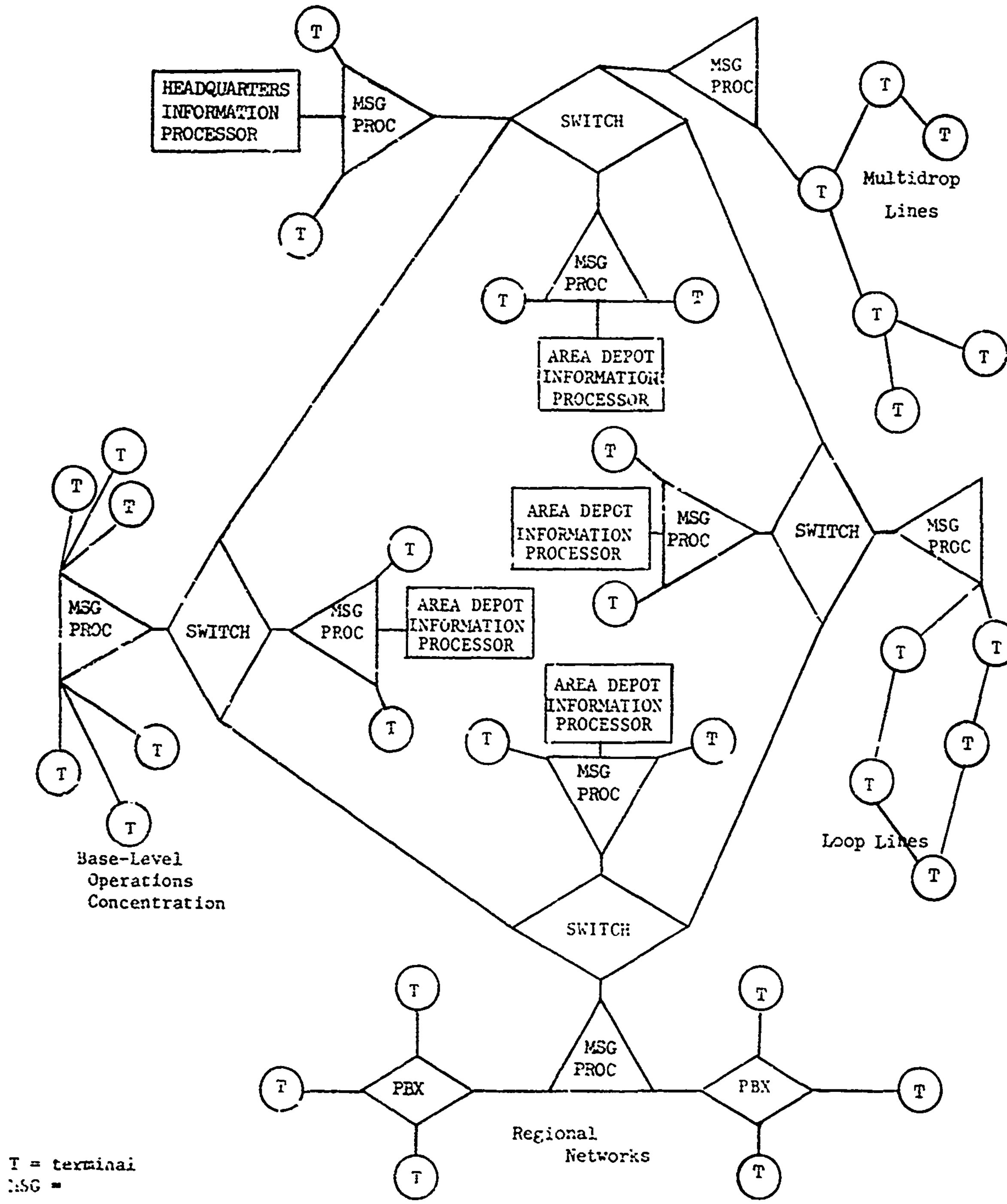


Figure 8. Logistics: A Typical Distributed Network of Interrelated Computers and Data Sets

2.1.3 Network Load and Traffic Patterns

In any command and control system, performance depends upon system capacity and the distributions of workloads. In some instances, fluctuations in workload due to the periodicity of tasks, time shifts, and changes in demand may be matched by the reallocation of resources such as load balancing across a number of computers or the alternate routing of messages. Some fluctuations are controlled through priorities and schedules. For instance, most computing facilities schedule short tasks requiring rapid turnaround during prime shifts and long production jobs during night and off-hours shifts. In a multipurpose command and control net, force-control tasks with high-activity periods may be alternated with background tasks of lower-priority command-support work.

For network analysis and evaluation, the flow of work in the system must be specified and performance requirements must be known. System performance criteria are usually complex, involving speed and volume of work, permissible error rates, reliability, security, continuity of operation, and measures of cost and cost-effectiveness. The nature and characteristics of the task to be performed influence the effectiveness of a system configuration. Some of the characteristics and performance criteria for computation and communication tasks are listed in Table 1. In addition, for network analysis, information must be available on the sources and destinations of jobs and messages over the total network. If the load fluctuates over time, fluctuations must also be represented in the workload specifications.

2.1.4 Operational Functions

Command and control systems are also characterized by their operational functions. These functions include planning, situation monitoring, resource monitoring and maintenance, warning and alerting, and execution and control.

Planning is necessary to almost every operation. It encompasses very-long-range strategic planning and policy formation and very-short-range planning for immediate operations and daily activities. It entails numerous functions in

TABLE 1

CHARACTERISTICS OF COMPUTATION AND COMMUNICATION TASKS

COMPUTATION TASKS		COMMUNICATION TASKS	
Item	Characteristic	Item	Characteristic
Job Length	Average Variance Distribution	Message Length	Average Variance Distribution Block Size within Message
Job Rates	Frequency of Performance Distribution	Message Rates	Average and Peak Arrivals Average and Peak Departures Arrival and Departure Distributions
Error Control	Permissible Error Rates Error Detection, Correction, and Control Procedures	Error Control	Permissible Error Rates Error Detection, Correction, and Control Procedures
Job Delay	Allowable Computing Time (Ignoring Computing Load and Terminal Use) Expected Throughput (Considering Load and Line Availability)	Message Delay	Allowable Switching Center and Line Times (Ignoring Message Load and Unavailability of Lines) Expected Delay (Considering Load and Line Availability)
Data Storage Requirements	Data Base Size File Sizes Input and Output Record Keeping and Accounting	Message Storage Requirements	For Record Keeping For Retrieval and Forwarding For Accounting and Statistical Data
Reliability	Acceptable Down Times, Recovery Times	Reliability	Acceptable Down Times, Partial or Complete Acceptable Recovery Times
Addressing and Routing	Acceptable Customers for Task Distribution of Results Security and Privacy Provisions Accounting Procedures	Addressing and Routing	Admissible Address Types (Multiple, Group, Single, Narrative) Average, Maximum Addresses in a Group Average Destinations for a Message Maximum Addresses in the System Alternate Route Computation Security Provisions
Job Type	Scientific, Business, Communications	Message Type	Narrative, Digital Data, Facsimile, Oral
Job Priority	Number and Functions of Priorities Job Recovery and Restart Procedures	Traffic Handling Procedures	Number of Priorities and Handling Conversion Requirements Network Interfaces Traffic Supervision of Overloads, Errors

in which computer and communication systems may assist the planner, as in statistical analyses of trends (forecasting), scheduling, allocation of resources, budgetary computations, and complex simulations of possible futures. It is most important in force-control systems, where precise contingency plans must exist to enable rapid response of massive and complex forces to emergency situations.

Situation monitoring is also important to force-control systems, especially at the tactical and operational levels. Immediate, up-to-the-minute information on the status of forces, the order of battle, and information concerning all aspects of an area of operations are important to successful operations. Communication systems for the rapid collection and dissemination of information and computers for the organization, analysis, and display of situational elements are essential to modern tactical operations and are only slightly less important in maintaining accurate, up-to-date information on all the other aspects of military affairs.

Resource monitoring and maintenance systems are similar to situation monitoring systems but employ special techniques to keep track of inventory (men, materials, and money); to ensure efficient acquisition, distribution, and transportation of resources; and to determine proper allocations.

Warning and alerting systems employ the results of planning and situation monitoring systems to compare an actual situation to a desired one and to report deviations.

Surveillance and detection systems frequently employ a large number of ordinary and exotic sensors, and special procedures are often included in other systems to permit them to react to the detection of threatening situations.

Execution and control systems vary from the direct guidance of vehicles to broad-gauge systems for the direction and management of large forces.

Despite the scope and scale of modern military command and control systems, the actual proportion of operational functions that has been automated is quite small. The degree of integration of these functions into a cohesively operating whole is smaller still. The potential for their further automation is tremendous, as is the promise in terms of speed, efficiency, and economy of operations.

2.1.5 Performance Characteristics

Of primary importance to system evaluation is performance. An information processing system has three major requirements: the flow of information through the network (communication capabilities), the processing of information (computing capabilities), and the supply of information (data base capabilities). Each of these has associated with it certain performance characteristics such as volume or throughput, performance times, reliability, accuracy, precision, and security. Hence, extremely complex relations exist among the factors contributing to the system construction and performance characteristics. Measurements of system performance such as vulnerability--the susceptibility of the system to performance degradation due to damage or failure--reflect the interrelation of performance characteristics and other factors. For instance, weights may be attached to certain items of information or to certain processes or interchanges, so that the loss of the capability may be judged of little (or great) importance. Weights may also be attached to various performance characteristics, as very great importance would be attached to rapid response times in operational-level force control systems. The factors contributing to the achievements of a particular performance level are also complex--as, for instance, the probability of a computation's not being performed, for example, is a composite of the failure probability of the hardware, software, and information components.

2.1.6 Sample Network Selection

In its mid-range plans, DoD already forecasts its needs and its allowable budgets for data transmission and processing in broad terms. Knowledge of such plans gives direction to the investigation but does not provide the specific traffic,

configurational, operational, and performance characteristics that must be planned for.

In lieu of a complete evaluation of DoD's computation and communication needs for now and for the foreseeable future, a representative system has been selected for study. This system may be expected to provide realistic operating concepts and guidance for the analytic (and/or simulation) models and typical processing and transmission loads and traffic patterns for trade-off and replacement studies. It must also:

- Be representative of the systems supporting major command missions and operational functions.
- Have a high probability of having available relatively accurate performance, load, and traffic-pattern requirements data.
- Have available resource personnel and other data sources to provide operational concepts and advise the project upon the reasonableness of assumptions and analytic results.

The complex of performance characteristics, functional operations, and operational missions that characterize military command and control systems is summarized in Figure 9. For the greatest scientific precision and generality of results, representatives of each of the classification categories should be studied. Long range, the Project hopes to study a broad sample of these systems. Short range, the Project has selected a single system, the Marine Manpower Management System (MMS), to be used in validating initial network analysis formulations and in studying simple trade offs. This system and the results of the validation effort are described in Section 4.

2.2 MODELING COMPUTATION AND COMMUNICATION NETWORKS

2.2.1 The Necessity of Modeling

To grasp the totality of interrelations and anticipate the impacts of events in one part of a complex network upon another portion is beyond unaided human

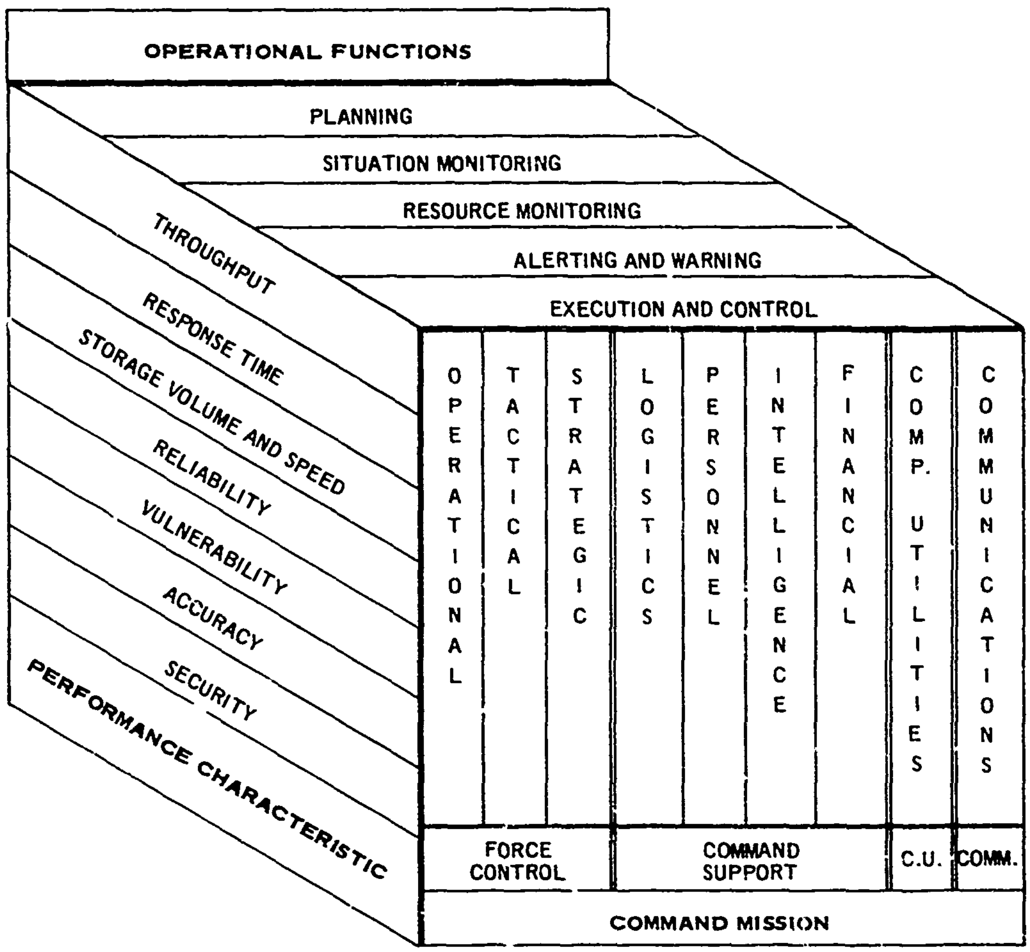


Figure 9. A Taxonomy of Command and Control Systems.

capabilities. To understand network behavior, some means of representation of network topology and traffic flow is required short of actually constructing a network and running practical experiments. A mathematical or logical model, simulating the operations of a network, will enable us to evaluate the impact of changing any characteristic of network construction, traffic flow, and performance.

' simulation model can:

- Provide a methodology for describing and analyzing the topological structure of the network.
- Determine optimal network flows and minimum-cost networks.
- Help in the investigation of network performance criteria such as response time, reliability, and vulnerability.
- Evaluate topological, performance, and operational alternatives for cost-effective trade-offs.
- Predict the impacts of technological innovation and component-replacement policies upon the cost effectiveness of a network.

There are several approaches to simulation, any of which may be used for network simulation depending upon the problem under investigation. In general, these are:

- Analytic Simulation--If a more or less steady-state or expected-value (mean) level of operations can be assumed, an analytic representation of the network that delivers a unique solution to a set of inputs can be used. Problems of maximum flow, minimum cost, and vulnerability yield readily to such modeling.
- Continuous and Discrete Simulation--If the flux and flow of temporally arranged events is the principal concern, many iterations of the analytic model, or a model that traces all events step by step through system modules, can yield a trace of changing performance with changing events. Problems involving sets of events, such as priority classes that do not yield readily to an analytic formulation, may also be attacked.

- Stochastic Simulation--If the functional relations evaluated in the model depend upon chance parameters, predetermined and expected-value parameters may be abandoned in favor of values randomly selected from appropriate event distributions. Under these conditions, many iterations of the model may be required to obtain a coherent picture of network behavior.

In CACTOS, we have already used the first and are developing means to use the others. Computation and communication network simulation involves the interface of communication-network flow analysis, largely centered in network channel behavior, with computer facility operation analysis, which inserts considerations of node behavior into the network simulation problem. The bulk of the work that has been done on transportation nets (which include communication, vehicular, and materials flow analyses and resource allocation problems) has tended to ignore node behavior beyond assigning a standard (or individual) delay time to nodes. (See Ford and Fulkerson, 1962; Kleinrock, 1964; and Frank and Frisch, 1971.) Similarly, the modeling done on computers and computing facilities has tended to give only slight attention to the fine evaluation of channel characteristics. (See Neilsen, 1970; Ihrer, 1967; and Schneidewind, 1966.) Solving the problems inherent in finding cost-effective trade-offs in telecommunication nets requires careful modeling of the computer/communication-net interaction.

Many problems exist for the simulation of networks and computers, including assumptions concerning the independence of events, the distribution of events, infinite queueing, and optimum routine algorithms. For instance, in simulations of multichannel, multicommodity, and multipriority systems, it is usually assumed that the flows of the members of a channel, commodity, or priority class are independent of one another even though in actual operations they may not be. This assumption greatly simplifies the queuing model.

Further, poisson and negative-exponential distributions are normally assumed for arrival rates and job sizes, also resulting in computational simplification.

Other distributions may be assumed or empirically derived and applied to traffic to determine their effects on estimated performance.

In analytic models, the impact of limited storage in both transmission pipelines and processing nodes is usually either ignored or simulated in terms of its impact on processing (including transmission) capacity. That is, infinite queueing is assumed. Hence, questions concerning distributed stores (e.g., data bases) and hierarchies of memories are not well answered.

Routing traffic may present problems when there are many alternative paths that could be followed between two distant points. Fixed-path routing is often assumed because determining the shortest, or most reliable, path to follow is relatively easy, but fixed-path routing is inadequate in the face of high load and limited line capacity. Allocation of traffic to alternative routes is also relatively easy if the paths are independent, but not when they form a web of interconnected branches with varying capacities.

Although the CACTOS project seeks to tie computation and communication network simulations together, initial models have been based upon the work already done in the field. Hence, assumptions have been made that have normally been made in graph and network analysis and in queueing theory. At the gross level of analysis anticipated for preliminary investigations, such models should prove adequate. Even so there are many challenging problems to be solved. In the future, as discrete and stochastic simulation models are developed, more stringent assumptions may have to be made.

2.2.2 Parametric Data Requirements

Although many network analyses consider only one or two attributes of links and nodes, there are potentially a great many such characteristics that can be associated with a net that could influence performance.

2.2.2.1 Performance Characteristics

Since the behavior of networks is one of our primary concerns, the kinds of performances evaluated will dictate many of the network attributes that must be considered in an analysis.

- Throughput Volume--To obtain a maximum flow through the network requires that the best arrangement of links and nodes and the best allocation of computing and communication capacity be identified. The primary attributes here are capacity, cost, distance, and traffic measures.
- Response Time--To obtain an acceptable average speed of response, the attributes to be considered are much the same as for throughput volume evaluation, but capacities and costs may be permitted to grow until the minimum amount is found that will obtain a required response to a stated load. Of course, optimally effective network configurations are still sought.
- Reliability and Vulnerability--If continuity of operation is the prime concern, the model must offer evaluations of the combined reliability of complexly interconnected nodes and links. While the evaluation of exact expressions for network reliability is complicated by the large number of terms that must be calculated, approximations may sometimes be more readily found. Methods exist for finding minimum articulation levels (the number of nodes and/or links that must be "destroyed" or down in order to break a network in two) and a measure of vulnerability (for evaluating the impact of the loss of a node or link on traffic flow). In the simplest case, reliability expressed as a proportion of downtime and reduced capacity may be used to degrade the average performance or capacity of the processor rather than evaluate reliability or vulnerability as objective functions.

- Accuracy--Much effort may be spent on the detection and correction of data-transmission and processing errors in a system if great accuracy is essential. As in the case of reliability, exact accuracy indices are difficult to derive, and the impact of accuracy levels on throughput and response time may be more easily assessed. For both reliability (failure rate and recovery times) and accuracy (error rates and regeneration times), the model could simulate the factors contributing to failure and malfunction and the strategies employed to compensate for them in considerable detail. However, unless reliability or accuracy are to be intensively studied, less stringent analyses are usually performed.
- Security--Military systems normally have rather high security performance requirements. Values such as the probability of penetration, or the probability of interception, and the minimum (or mean) time to decrypt messages may be used as security values associated with channels, processors, and data. Again, evaluation of the security of a network involves many of the same problems as do evaluations of reliability and accuracy. However, obtaining additional security for a system usually increases costs more than it reduces throughput or responsiveness although extra processing may be involved.

2.2.2.2 Topological Considerations

Basically, a network consists of nodes and links. Nodes usually have specific locations, although in the case of ground, water, air, and spaceborne vehicles, the location may only be momentary. Links usually have lengths, subject again to the constraints of mobility. For flow analyses, both links and nodes must have processing capabilities and capacities, which may include limitations on direction of flow and degree of storage. For general analyses, capacity values may be assigned directly to the nodes and links; in the evaluation of alternative configurations composed of actual or proposed transmitters and processors, it is easier to assign sets of processors to the nodes and links, because of the plethora of possible devices and the multiplicity of their characteristics and functions.

2.2.2.3 Processor Characteristics

Both data transmission channels and data processing facilities are complex devices. Transmission lines may contain transceiving equipment that does some transformations on the data; computing equipment operations involve many transmissions of data to and from memory and I/O devices. If close comparisons of alternative configurations and alternative distributions of capacity are to be made, the characteristics associated with a processing device must include not only capacity, reliability, and accuracy, but the details of the structure of data, the functions performed, the compatibility with other devices and operations, and other operating details.

Some computation and communication processor characteristics that might be useful to a teleprocessing system model were listed in Table 1. People and programs are not included.

Two "processors" frequently not modeled in communication and computation simulations are the human operators and the processor software. Shinners (1967) summarizes some of the work that has been done on modeling the human transfer function, but, in general, little statistical data concerning operator behavior has been compiled for data processing operations. Sackman (1967) has summarized the rather sparse information concerning user behavior at on-line terminals, but the time lags associated with human processing in the computing or communication facility (tape and disc mounts, dismounts and transports, job setup and tear-down times, etc.) have either not been systematically studied or vary so greatly from situation to situation that model formulations are difficult

Software simulation modeling has received some attention both from a costing and an operational impact view. SCERT (Ihrer, 1967), for instance, includes software parameters in its job operating model. Modeling of various operating systems, however, often contain uncertain estimates of operating system overhead costs. Estrin and Kleinrock (1967) have summarized many of the models used for time-sharing system simulation with different queuing disciplines. Less precise formulations are usually made of applications software.

2.2.2.4 Traffic Characteristics

Most of the characteristics of messages and computation jobs are given in Table 2. These characteristics reveal some indication of the flow of traffic and the interrelation of jobs. However, they do not fully reveal the shifts that occur in job and message mix and volume with the time of day or the way traffic queues form, storage fills up, and bottlenecks occur. When models are formulated, decisions must be made about the handling of traffic control for communications and job operating overhead for computations. These events tend to be small in size but occur quite frequently, so that including them in statistical representations of traffic can be misleading. A separate formulation to evaluate traffic control effects before inclusion in run results should be made.

For modeling purposes, traffic simulation can be quite gross, but for others it must be quite detailed. A computing task may involve inputs and outputs from several terminals, auxiliary storage devices, and communication lines. Inputs arrive over several parallel channels operating at different speeds and using different codings, record sizes, and storage structures. Queueing, buffering, and priority requirements may differ, as may those for security and accuracy. The performance of the system may be greatly influenced by the degrees of I/O overlap attained, by the efficiency with which the CPU is used in processing jobs from several queues, and by the storage-allocation efficiency. Message processing simulation seems somewhat simpler, but encounters many of the same problems in handling alternative routings, complex multiplexing, and switching.

At a minimum, a network model must consider job and message arrival rates, sizes, sources, and destinations. Priorities, parallel operations, and job and message types may be added, and complex shifts in traffic composition in time and space may be simulated by appropriate sets of inputs. In performing experiments, much manipulation of traffic characteristics may be necessary to create particular effects.

TABLE 2

CHARACTERISTICS OF COMPUTATION AND COMMUNICATION DEVICES

Central Processing Units

Word size	Reliability
Instruction rate	Accuracy
Immediate storage volume	Security rating
Immediate storage transfer rate	Cost
Number of I/O channels	Pointers to peripheral units and communication units

Information Storage Units

Storage type	Reliability
Storage register size	Accuracy
Number of registers per unit	Security
Number of permissible storage units	Cost
Average access time	Pointers to compatible processors and other devices

Terminal Processors

Terminal type	Reliability
Data Structure(s)	Accuracy
Temporary storage	Security
Access time	Cost
Transfer ratio	Pointers to compatible devices
Operating functions	

Terminal and Transmission Control Units

Type of control exercised	Access time (transmission)
Device(s) controlled	Transfer rate (terminal)
Number of terminal channels	Transfer rate (transmission)
Number of transmission channels	Reliability
Type of transmission	Accuracy
Data structure	Security
Available storage	Cost
Access time (terminals)	Pointers to compatible terminals and channels

Communication Channels

Type of media	Reliability
Type of modulation	Accuracy
Type of multiplexing	Security
Number of subchannels	Cost
Transmission rates	Pointers to compatible control devices

2.2.3 Network Analysis Techniques

The basic contributions to the study of networks have come from the study of electronic communications systems and operations research models for economics and distribution systems. Mathematical tools include graph theory, combinatorics, probability theory and statistics, mathematical programming, and queueing theory. The basic problems that are considered have to do with maximizing flows, minimizing costs, handling multiple queues (terminals and commodities), assessing connectivity and vulnerability, and evaluating time delays and throughput. How these problems are handled depends upon whether the analyst is seeking a deterministic or a probabilistic solution.

2.2.3.1 Maximum Flow

The aim of many network flow analyses is to maximize the flow, subject to a set of constraints. The constraints may be specified line or node capacities, costs, or response times. The analytic algorithm used is based upon the max-flow/min-cut theorem, which states that the maximum flow of value from a source to a terminal is equal to the minimum cut capacity of all cuts separating the source and terminal, a cut being any set of links (or node chain or node-and-link set) connecting the source and terminal whose elimination would separate the source and terminal. The cut capacity is the sum of the capacities of the branches of the cut.

The analysis may be via combinatoric, graph-theoretic, or linear-programming methods. The most widely used approach is the Ford and Fulkerson Out-of-Kilter Algorithm--a combinatoric solution that applies to a variety of problems including multiple-commodity and multiple-terminal problems. For instance, the most reliable path through a net may be found by minimizing the sum of the logs of the probabilities of failure (or percent of downtime), which is equivalent to a product of reliabilities.

2.2.3.2 Minimum Cost

Maximum-flow problems are normally considered without regard to the constraints of cost. Minimum-cost problems are concerned with finding an economic solution

to flow problems such as shortest paths (or a ranking of paths), least-distance paths, shortest trees, minimum dollar cost, and specified reliability. Best-location problems also fall in this category where average path length, average transmission time, or average construction cost of a switching center or some other facility is the cost to be minimized.

Conceptually, the technique used to resolve minimum-cost problems is quite similar to the Out-of-Kilter Algorithm, which assumes constant costs. The max-flow/min-cut theorem is used in finding paths and trees to which convex cost functions may be applied. Flow is then increased along the least-cost-augmentation path until the balance point is reached.

Finding least-cost locations (where "cost" may be distance, time, transportation expense, probability of failure, etc.) is very similar. The problem is not altogether simple since some paths may be one directional so that the cost center of a complex network may not be intuitively obvious. Further, in designing a network, a set of local "medians" may be sought, in addition to an absolute center, where switching centers would be located to minimize the total line length (or total response time or costs) in the system. Procedures exist for considering established vertices or for determining whether there are non-node locations on the paths that yield a better solution.

Other procedures may be used for finding the minimum-cost network to satisfy terminal-capacity requirements, to determine least-cost improvements, and to consider the impact of losses and gains (e.g., error, noise, and signal enhancements) in the branches of a network. While exact solutions of some of these problems rest on computationally extensive linear programming techniques, shorter graphic procedures are available for many conditions.

2.2.3.3 Connectivity and Vulnerability

The vulnerability of a network (the potential loss of communication due to failure or deliberate destruction) is a most important consideration for military

systems. The analysis of vulnerability depends upon the criterion chosen for "destruction" or "loss of communication." Some possible criteria are:

- Loss of Connectivity--The net is divided into two or more subnets by the removal of one or more nodes or links. The number of links or nodes that must be destroyed or fail before separation occurs is an index of vulnerability.
- Loss of Routes--Failure, destruction, or overload results in there being no direct path available between two or more sets of specified vertices.
- Degradation--Less than some specified number of nodes remain in service, or residual capacity falls below some minimally required amount.
- Minimum Cost--The shortest path, least distance, residual reliability (probability of failure), response time, or some other cost measure is greater than some specified value.

Loss of connectivity is the simplest and most obvious of these criteria and, since cut-set algorithms are available in the standard analytic procedures, one that is readily calculated. If a mixed cut-set (the minimum set of either or both nodes and links) is desired, some further analysis is required. In network construction, methods for maximizing the smallest cut-set and for attaining specified levels of branch redundancy are also available in achieving relative invulnerability.

Minimum-cost methods are somewhat more difficult, and exact solutions may not be attainable. The problem investigated is the probability of a call's being made (i.e., of there being an available connection) under conditions of random failure (or unavailability for other reasons), and this requires more complex evaluations.

2.2.3.4 Time Delay

Time delay (network response time) introduces the notion of queues and storage into network analyses. That is, it is assumed that units of flow arrive at a processing station in a stochastic fashion rather than in a steady-state flow

and that irregularities in arrival will result in some units' being required to wait until the station is free. Storage (memory) must be assumed to hold the queue of waiting units.

Problems here focus on the time required to process a job or message, upon routing problems in obtaining an optimal flow, and upon determining optimal storage and processing capacities to satisfy flow or cost requirements. These are the very problems of greatest interest in digital data processing and transmission systems. The basic formulation of the problem is quite simple, but problem solutions rapidly become complex as the simulations of traffic patterns and processor idiosyncracies become more faithful and detailed.

2.2.3.5 Throughput

Computer throughput may be taken as an example of the simulation of traffic patterns and processor idiosyncracies. The job-processing capacity at a computation node does not depend entirely upon the raw computing power of the central processing unit but upon the characteristics of the job being processed and upon how well the job is set up to take advantage of the capabilities of the computer. The simplest assumptions to be made are that transmission and computation are either completely sequentially dependent (no computation is done until input is complete, nor is output made until processing is over) or completely independent (computation and transmission occur without regard for the state of completion of the other). Except in relatively primitive cases, these are not valid assumptions. Most computers have a number of I/O channels serving a variety of storage and I/O peripherals. Jobs are set up to perform one or more computations and one or more data transfers in parallel, depending upon the number of processors and channels available and the number of independent tasks in the processing queues. However, no matter how cleverly the procedure is set up, there will be some mismatches among parallel processes. The immediate storage capacity of the computer, which is used to buffer among competing units and hold work queues, is limited. This constrains the degree and efficiency of the buffering process.

In practice, one may express the complex job-processing arrangement as a subnet and its network flow or one may express analytically the impact that job characteristics have upon the processing power of the computer. For fine investigation of computer facilities, computer operating systems, and program and data structure, the simulation is desirable. For gross network analyses, estimates of restrictions on processing power are probably adequate.

2.2.4 Interactive Processing Requirements

If a network analysis and simulation model is to be of maximum use to the analyst, it should permit on-line, conversational interaction, not only performing the selected network analysis functions but communicating with the user easily and efficiently. Desirable features in the user interface are:

- Man-Machine Dialog--In on-line use, a constant interplay of question-answer, response-reply dialog should be maintained to guide and assist the user in using the analysis program.
- Alternate Input/Output Modes--Both on-line and off-line input/output modes are required; on-line to handle small problems and modifications and off-line to handle voluminous inputs and outputs.
- On-Line Editing and Modification--To correct input errors and to adjust model parameters over a large range of alternatives during exploratory investigations, extensive on-line editing and modification capabilities are needed.
- Storage Capabilities--Data bases are required to hold files of pertinent information such as processor characteristics. Working storage is required to hold sets of experimental conditions between sessions and between successive exploratory trials.
- Query Capability--The user needs to be able to interrogate data files of work in process, in temporary storage, and in data base files for feedback assurance on correctness of inputs, for reminders of previous conditions and results, and to retrieve appropriate information for further work and decisions.

- Function Library--In addition to selecting the sorts of network analysis functions to be performed during a particular investigation, the user needs capabilities for desk calculation, statistical analysis, and a variety of algebraic functions to assist him in evaluating results and making decisions.
- Graphic Displays--Graphic displays of networks and plots of results are highly desirable as design aids, as feedback of results, and as evaluations of interrelationships.

2.2.4.1 Man-Machine Dialog

Interaction between the analyst and the model must be sufficiently "English-like" to be readily understood by the analyst, but, since the bulk of the information exchanged is numeric data (largely matrices or tables), extensive conversational exchanges are not necessary. Two input procedures seem desirable: a "help" mode to lead the analyst through the intricacies of the modeling process and a terse, efficient exchange to avoid the inefficiencies of verbose queries and responses. In view of the potential error rates for human operators, the interaction should be as "forgiving" as possible and considerable input checking should be done.

2.2.4.2 On-Line Editing

For easy exploration over a variety of conditions and easy correction of input errors, the model should be designed to permit the analyst to modify values readily in all tables and matrices and to iterate readily through individual analytic junctions or complete analyses.

The analyst should have the capability to:

- Input a complete matrix, or any row, column, or value of a matrix.
- Set a matrix, or a row or column of a matrix, to a specified value--i.e., a constant.
- Change a single, specified value in a matrix or table.

- Modify a matrix, or a row, column, or value of a matrix, by a mathematical operation (add, subtract, multiply or divide by a scalar, vector, matrix or function). Input values should be expressible as integers, decimals, or exponentials.
- Set a portion of a matrix (a row, column, or single value) to a specified value or specified sequence of values and set the remainder of the matrix or matrix portion to a constant value.
- To select in an arbitrary sequence the tables to be input, modified, or queried.

2.2.4.3 Alternate Input Modes

Basically, the model should be capable of operating in on-line, batched, or mixed modes. In mixed-mode operation, the analyst should be able to specify input values on-line or to specify the data base or the input device from which the input values are to be taken. Outputs should be directable to the user's console, to an on-line printer or display, or to storage for delayed output or for selective query.

2.2.4.4 Storage Capabilities

Much of the information concerning technology, traffic flows, and network construction is fixed independently of a particular problem or run. Data bases may be created to encompass technological data and analysis in progress, and analytic results may be stored for either retrieval or use in a problem run.

2.2.4.5 Queries

There are a great many queries that the network design and analyst would like to ask of the model. Some of these are straightforward requests for the display of input parameters and results (e.g., the ij^{th} entry of a matrix). Others would require retrieval from a data base (e.g., the processor characteristics of a CDC 6600 computer). Still others might require model computations (e.g., the shortest path between points i and j and its load).

It is desirable to have a relatively flexible query language, one that would permit either relatively formal statements in a terse symbolic form or more informal, English-like expressions. Terse inquiries are convenient for rapid professional work; more informal language is desirable for users and for exploratory interrogations.

2.'.4.6 Function Library

In setting up any simulation model, there is a great deal of preparatory work that must be done, largely involving the derivation of statistical parameters (means, standard deviations, correlations) and simple arithmetic calculations. In interpreting results, similar questions may be asked about subsets of values and particular relationships. Functions required include:

- A "desk calculator" that will accept numeric statements for evaluation or algebraic formulations for repetitious computations.
- Statistical routines including means, standard deviations, and correlations under varying assumptions and distributions.
- Simple algebraic functions such as logs, roots, maximums and minimums, trigonometric functions, and matrix operations.
- Curve-fitting routines for graphic plots.

2.2.4.7 Graphic Displays

Graphs of performance variables plotted either against other performance variables or against model parameters give the analyst an integrated and comprehensive picture of network behavior and are valuable in understanding the behavior and in advancing further hypotheses.

For the designer, graphs, such as CRT plots of the networks being designed into which the designer could insert links, nodes, and specific values, are powerful design tools. If such graphs were fully interactive, they would afford the designer immediate feedback on the results of his actions.

2.2.5 The Prototype Analytic Model

SDC has developed an on-line network analysis program entitled DESIGNET (Citrenbaum, 1971) that can be applied to a variety of problems (communications systems, power distribution systems, allocation and scheduling over networks of repairmen, installers, policemen, and tactical air support, etc.). The initial version of this program has rather limited network analysis capabilities, being oriented toward evaluation of delay time rather than optimization of flow, costs, or construction details. For example, it includes a modified version of the Out-of-Kilter Algorithm that will compute only optimal fixed-route flow rather than a maximized network flow. However, the program incorporates both computation and communication response time analyses, and an evaluation of the impact of job characteristics (a throughput model) on computer performance is available as an option (Cady, 1971).

DESIGNET is programmed in FORTRAN IV and operates under the ADEPT or TS/DMS operating systems in on-line or semibatch modes. The initial program has no data base capabilities for advance storage and retrieval of descriptions of processors, lines, networks, or traffic, but the operating system permits saving of intermediate results (the /SAVE function) during exploratory investigations. The operating systems used at SDC also have a desk calculator capability (/TRIP) and other facilities to assist the on-line analyst.

2.2.5.1 Construction Techniques

The initial input to the analytic program is the node-link structure of the network to be analyzed. The program allows two alternative means of network specification, either by listing all the individual links in the net or by listing the node interconnections. For example, consider the network shown in Figure 10.

Figure 10. A Sample Network.

A link specification of this network consists of the seven links as shown in Figure 11A. A node interconnection specification consists of a listing of all nodes connected to node $\underline{i}$ as prompted by the computer (Figure 11B). If all lines are duplex (i.e., bidirectional), then the node interconnection specification can be simplified to that of Figure 11C.

A	B	C
1-2	1-2,3	1-2,3
1-3	2-1,3,4	2-3,4
2-3	3-1,2	3-(blank)
2-4	4-2,5,6	4-5,6
4-5	5-4,6	5-6
5-6	6-4,5	6-(blank)
4-6		

Figure 11. Three Equivalent Specifications of Network Structure.

2.2.5.2 Topological Network Characteristics

Research into graph theory (Berge, 1962; Ore, 1962) and discussions with applications engineers to identify the network parameters of interest led to the description of a network structure in terms of ten parameters:

- P_1: Number of nodes in the net.
- P_2: Number of links in the net.
- P_3: Ratio of the number of links to the number of nodes.
- P_4: Ratio of the number of links to the number of links in a fully connected network--A measure of network "fullness."
- P_5: Variance of link-to-node ratio--A measure of the clumpiness of the net.
- P_6: Minimum node connectivity--The fewest number of links attached to any node in the network.
- P_7: Network radius--The shortest path (in terms of links traversed) from the most central node in the net to the node most distant from it.

P_8: Network diameter--The shortest path (in terms of links traversed) between the two most distant nodes.

P_9: Articulation--A measure of network vulnerability calculated for both nodes and links. Consists of.

- P_{9a}: Node articulation level--The fewest number of nodes which, if deleted, would break the network into at least two non-communicating subnets.
- P_{9b}: Link articulation level--The fewest number of links which, if deleted, would break the network into at least two non-communicating subnets.
- P_{9c}: Articulation nodes of level n--Those minimal sets of n nodes which, if deleted, would break the network into at least two noncommunicating subnets.
- P_{9d}: Articulation links of level m--Those minimal sets of m links which, if deleted, would break the network into at least two noncommunicating subnets.

P_{10}: Closed loops--A measure of network redundancy and available round trips, consisting of:

- P_{10a}: The number of unique closed loops in the network.
- P_{10b}: A listing of node sets contained in all such unique closed loops having more than n nodes, where n is specified on-line by the user.

For the network shown in Figure 10, the parameters are:

$P_1 = 6$

$P_2 = 7$

$P_3 = 2.33$ (for duplex lines)

$P_4 = .46$

$P_5 = .22$

$P_6 = 2$

$P_7 = 2$

$P_8 = 3$

$P_{9a} = 1$

$P_{9b} = 1$

$P_{9c} = 2,4$

$P_{9d} = 2\text{-}4$

$P_{10a} = 4$

$$P_{10b} = \begin{cases} 1,2,3 \\ 3,2,1 \\ 4,5,6 \\ 6,5,4 \end{cases}$$

2.2.5.3 Computation-Communication Network Analysis

Description of the model for computation-communication analysis can best be made by identifying four model characteristics: (1) inherent assumptions, (2) necessary user inputs, (3) modification capability, and (4) model outputs.

(1) Assumptions--The assumptions upon which the mathematical analyses are based include:

- Job and message interarrival times are independent of job and message lengths. (The independence assumption is basic to the queueing model.)
- Job and message interarrival times and lengths have poisson distributions.
- Communication nodes have infinite message buffer size (i.e., node storage permits infinite queues) and processing (transfer) capability.
- A constant delay time for internal message processing is assigned each node.

(2) Inputs--There are nine primary inputs to the model plus eight more for computer throughput evaluation. The primary inputs are:

- Network Structure. Nodes and links are specified by either link pairs or node associations.
- Distance Matrix. The distances between nodes may be set to a constant, may be input via a matrix of distances between all node pairs, or may be calculated from a list of input node locations in latitude and longitude.
- Job Arrival Rates. Job arrival rates may be set to a constant or specified as a job arrival matrix.
- Message Arrival Rates. Message arrivals may be set to a constant or input as a matrix or not specified. If not specified, the model assumes message arrivals are the same as job arrivals.
- Job Size. Job size in megabits of computer processing may be set to a constant mean job size or input as a matrix of mean job sizes from node i to be processed at node j.

- Message Size. Message size in bits per message sent from node i to node j may be set to a constant or input as a matrix of mean message sizes between nodes.
- Job Processing Rate. The job processing rate is expressed as the number of megabits a computer is capable of modifying per second and may be input as a constant, as a vector of rates, or assigned to each node under the guidance of the model. The job processing rates used by the model may be adjusted as a result of a throughput analysis.
- Channel Capacity. The capacity of each link in the network in kilobits transmitted per second may be set to a constant or input as a matrix of capacities between adjacent nodes. Alternatively, a total channel capacity may be specified for the network and the model will obtain a "square-root" channel capacity assignment for each line in the network on the basis of channel usage.
- Packet Size. A standard packet size in bits per packet may be specified. The model will divide the messages into packets and adjust its internal message-arrival and message-size matrices to reflect this information during computations.

For throughput analysis, several more input parameters must be specified:

- Computation Time. The fraction of a job's total time spent in computation.
- Average Record Size (in bytes). Records are assumed to be stored sequentially in secondary (disc) storage.
- Core Memory Size (in bytes).
- Access Time. The mean I/O unit access time in milliseconds.
- Transfer Rate. The I/O unit transfer rate in bytes per millisecond.

- Instruction Rate. The number of instructions per second that the computer is capable of performing (normally taken as the total possible additions per second, including fetch, add, and store).
- Word Size. The word size in bits of the computer.
- Computation Overlap. The fraction of I/O time that occurs simultaneously with computation.
- I/O Overlap. The fraction of I/O time that overlaps with other I/O in a multichannel system.

Since mean jobs must be described for each computation node in the network, each of these input parameters results in a vector of values. As with other sets of values, the vector may be set to constants. (All jobs performed and computer systems used are, on the average, similar. That is, job size, percent overlap, core memory size, record size, etc., are the same for jobs at all nodes.) Or, the values may be input differently for each computer.

(3) Modification Capability--During a modification phase, input matrices may be altered by a factor (multiplication or division) or new values may be substituted for existing entries. Modification may be by individual entries, or an entire row or column may be set to a constant or a vector sequence of values given. If an incomplete vector is given, all remaining entries in the row or column will be set to the last value specified.

(4) Outputs--The outputs of the network analysis program are: a network description, performance characteristics, link and node summaries, and on-line graphics.

- The network description consists of values for the 10 parameters.
- The performance characteristics include:

  Total Network Traffic--The number of messages per unit time moving through the net, excluding acknowledgments.

  Average Path Length--The mean number of links traversed by messages in the net.

Mean Communication Response Time (T_m)--The mean delay time of a message through the net, including transmission times and times in queues.

Mean Computation Response Time (T_c)--The mean delay time to process a job at a computer in the network, including both time in the job queue and computer processing time.

Total Response Time--The mean time to complete a job in the network, assuming jobs originating at location i are transmitted to location j for processing and results are returned to location i (i.e., $T_c + 2T_m$).

Total Cost--The approximate monthly lease cost for both computers and communication lines of the specified capacities.

Link and node summaries (optimal requests may be made for either, both, or none). Contain:

Job or message arrivals

Computer or channel capacity

Transmission or processing time

Time in queue

Computer or link leased cost per month

- The on-line graphics. Will output select performance- or network-characteristic parameters plotted against one another over specified ranges of the variables.

2.2.5.4 Communication Network Analysis

A fixed routing procedure (optimizing flow over links) for all messages passing through the network is calculated by a modified version of the Out-of-Kilter Algorithm. While fixed routing is not optimal in real-world applications, it is close enough to yield meaningful analysis. Using fixed routing and the assumption above, the mean queueing delay T_{a_i} on the i^{th} link is given by Equation 1.

$$T_{a_i} = \frac{\lambda_i/\mu_i C_i}{\mu_i C_i - \lambda_i} \qquad (1)$$

where: λ_i is the traffic per unit time on the i^{th} link

$\frac{1}{\mu_i}$ is the mean message size on the i^{th} link (1)

C_i is the capacity of link i

Similarly, the mean transmission and propagation delay T_{b_i} on the i^{th} link for a given message is given by Equation 2.

$$T_{b_i} = \frac{1}{\mu_i' C_i} + \frac{L_i}{\nu} + k \qquad (2)$$

where: L_i is the length of link i

ν is the propagation speed of the message through the channel media

k is the constant delay for message processing at the destination node

μ_i' is the mean "real" message size on the i^{th} link (without averaging in the acknowledgments)*

Kleinrock (1964) has shown that the total mean communication response time averaged over the entire network can be expressed by Equation 3

$$T_{CM} = \sum_i \frac{\lambda_i}{\gamma} (T_{a_i} + T_{b_i}) + k \qquad (3)$$

where: γ is the total network input data rate.

Equation 3 weights the delay on channel C_i with the traffic λ_i carried on that channel.

---

*If no acknowledgments are present, $\mu_i' = \mu_i$, and the following algebraic reduction occurs:

$$\frac{\lambda_i/\mu_i C_i}{\mu_i C_i - \lambda_i} + \frac{1}{\mu_i C_i} = \frac{1}{\mu_i C_i - \lambda_i}$$

If the individual channel capacities are to be selected to minimize the response time T subject to a fixed cost constraint, then Equation 4 gives the capacity of channel i under the assumption that channel costs functions are linear.

$$C_i = \frac{\lambda_i}{\mu_i} + \left(\frac{\sum_i d_i C_i - \sum_i \frac{\lambda_i d_i}{\mu_i}}{d_i} \right) \left(\frac{\sqrt{\lambda_i d_i / \mu_i}}{\sum_j \sqrt{\lambda_j d_j / \mu_j}} \right) \tag{4}$$

where d_i is the dollar cost per unit of capacity of channel i.

Under the further assumption that all channels cost the same regardless of length, this reduces to Equation 5.

$$C_i = \frac{\lambda_i}{\mu_i} + \left(\sum_i C_i - \sum \lambda_i / \mu_i \right) \left(\frac{\sqrt{\lambda_i / \mu_i}}{\sum_j \sqrt{\lambda_j / \mu_j}} \right) \tag{5}$$

Equation 5 is used by the model to give a good approximation to choice of channel size for minimizing network response time.

The monthly lease cost of each channel is presently obtained via a formula derived from least-squares fit to Telepak data. In the case where a duplex line has been specified and optimal capacity size computed, the leased-line cost is to be the cost of the larger of the two directional lines. In near-term extensions of the model, computed leased cost will be replaced with actual lease costs of various line types and sizes.

2.2.5.5 Computation Analysis

The computation analysis initially determines the number of jobs to be processed per unit time at each computing center and determines the feasibility of accomplishing desired computation in terms of available processing power. If computation is deemed feasible, the delays due to queueing and to processing at each node are calculated and used to compute the mean computation response time.

Calculation of the number of jobs to be processed at node i is a simple summation of the elements in column i of the job arrival rate matrix. Computer processing at node i can be accomplished only if jobs are processed at least as fastas they are arriving; that is, processing is feasible at node j if Equation 6 holds:

$$\sigma_j P_j - \theta_j \geq \emptyset \qquad (6)$$

where: θ_j = job arrival rate per unit time at node j

P_j = job processing rate per unit time at node j

$\frac{1}{\sigma_j}$ = average job size at node j, obtained from Equation 6a:

$$\left(\sigma_j = \frac{\theta_j}{\sum_i \frac{\theta_{ij}}{\sigma_{ij}}} \right) \qquad (6a)$$

The mean queueing delay for computation at node j is given by Equation 7 and the mean processing delay at node j by Equation 8.

$$T_{d_j} = \frac{\theta_j/\sigma_j P_j}{\sigma_j P_j - \theta_j} \qquad (7)$$

$$T_{e_j} = \frac{1}{\sigma_j P_j} \qquad (8)$$

The sum of the queueing and processing delays reduces to a simple relation as shown in Equation 9, and weighting this over the network inputs yields the mean computation response time in Equation 10.

$$T_{f_j} = T_{e_j} + T_{d_j} = \frac{1}{\sigma_j P_j - \theta_j} \qquad (9)$$

$$T_{CP} = \frac{\sum_j \theta_j T_{f_j}}{\sum_j \theta_j} \qquad (10)$$

To obtain the approximate monthly lease cost of each computing center, the model presently uses a least-squares fit to empirical data. As with communication cost data, this will be changed to allow actual lease costs of specified system configurations.

The total response time is calculated by multiplying twice the mean communication response time by the percentage of jobs processed at remote computing centers and adding this to the mean computation response time as shown in Equation 11.

$$T_T = 2ST_{cm} + T_{cp} \qquad (11)$$

where: S is the percentage of jobs processed at remote computers.

2.2.5.6 Throughput Analysis

Although the assumptions of infinite core memory and maximal computer processing power have the prime advantage of simplicity, they yield a rather naive representation of job processing. Differences and trade-offs among processing hardware, memory, I/O equipment, and job organization cannot be evaluated.

A throughput model was developed to evaluate overlapping I/O, finite core memory constraints, and transfer rate limitations (Cady, 1971). The model assumes infinite secondary storage and I/O channels--i.e., no queueing delays result from full discs, tapes, or drums, or from busy channels. A single processor at each node modifies one word of storage at a time--i.e., no multiprocessing or multiword modification is done. The number of I/O accesses is assumed to be proportional to available core storage--i.e., more records are read per access as core memory is increased. Computer innovations and operations such as streaming and parallel processing are ignored. Phenomena associated with operating systems (e.g., peculiarities of queueing disciplines such as round-robin or interrupt-driven systems) are also ignored. The formulati n of throughput assumes a standard computer (IBM 360/50 with 2314 disc packs and 500K bytes of core memory) against which the performance of other systems is compared.

The formula employed is:

$$P = \frac{5.4\ wpfm^2}{5.4\ fm^2 + 2.5\left(10^3\right)(1-f)v\left[wp\left(a + \frac{r}{x}\right) - 1.92\left(10^{-5}\right)\left(87.5 + \frac{r}{312}\right) mn^2\right]}$$

where:

P = the performance of the object computer (megabits modified per second).

f = the fraction of the job's total elapsed time spent in computation.

r = average record size in bytes.

m = core memory size in bytes.

a = mean I/O access time in milliseconds.

x = I/O unit transfer rate in bytes per millisecond.

p = instruction rate of the computer in instructions per second.

w = word size of the computer in bits.

n = fraction of the I/O time overlapped with CPU time.

v = the ratio of overlapped I/O to strictly sequential I/O.

2.2.5.7 Applications and Limitations

The prototype analytic model discussed here is under continuing development (not all of which need be directly applicable to communication nets). While the ultimate goal would be to develop a completely general computation-communication network analysis model, current capabilities are used only partially for communications. Some additional limitations are:

- All switching is assumed to be store-and-forward message switching although circuit-switching models are being considered.

- While fixed routing minimizes the number of links traversed by a message and is near optimal, it does not maximize total flow.

- While square-root channel capacity assignments yield good empirical results, they are not mathematically consistent with the channel cost equations.

- Using the number of bits modified per second as a measure of computing capacity is a poor approximation of the complexities of modeling a computer system. This situation is ameliorated somewhat by the throughput model, although this is admittedly still a relatively crude approximation.

- All transactions are of equal priority. That is, multiterminal, multicommodity and multiqueue models are not included. However, some attention is given to the differential treatment of long messages versus control signals (acknowledgments) in the communication model formulation.

- Distributions other than negative exponentials may better represent job and message traffic characteristics. However, these models yield computationally simple and apparently empirically good results. The poisson models must be used until contrary evidence (i.e., empirically established distributions) is determined.

Despite these limitations, the prototype model is applicable to a variety of studies. First, by directing attention toward the underlying characteristics of a network, it indicates areas in which additional analytic data should be gathered. Through the network description, the model provides insight into the topological peculiarities of the network. Measures such as link-to-node ratios, connectivity, articulations, radius, and diameter allow the designer to visualize beyond the real-world application and consider more penetrating

topological trade-offs. For instance, a decrease in network diameter (the shortest path between the two most distant nodes) is likely to result in a decrease in response time and error rates in the system.

Node and link articulation levels are useful in studying network vulnerability. Pinpointing vulnerable links and nodes, and a knowledge of actual, rather than modeled, conditions will help provide insight into operational optimization.

With the analytic model, the designer can use his experience and intuition to modify the network structure for improvements in network performance--response time or cost. By removing links and nodes, he can gain further insights into vulnerability and traffic bottlenecks. He may observe queueing delays at various locations via the individual link and node summaries and investigate cost distributions, channel capacity allocations, and other specific link and node behavior.

3. TECHNOLOGICAL ALTERNATIVES

In designing an information processing network of computers, communication channels, and control and display devices, the system engineer is faced with many structural and operational options. Design decisions are determined as much by the concept of operation held by the using organization as by the volume, nature, and traffic pattern of the computing, transmission, and information storage loads that are placed on the system. Fundamentally, however, we want decisions to be made on the basis of the relative cost-effectiveness of the design options, within the limitations of available funds.

Military information systems must meet high performance standards for reliability, continuity of operation, security of information, accuracy (freedom from errors), speed of response, and resistance to saturation. Although it is difficult to assign precise values to increases in performance capabilities, it is possible to formulate the cost and performance relations among performance requirements and design options and to determine the technical and economic trade-offs that exist among them.

Some of the trade-offs that have been postulated involve the economies of scale, specialization, topological configurations, integration of functions, improvements in quality, and incorporation of technological advance. Previous investigations (e.g., Sharpe, 1969) have provided some formulations of these economies for computation and communication systems. The ultimate aim of the CACTOS Project is to provide further formulation of these trade-offs, to incorporate them in network simulation models, and to examine their implications for military command and control systems. Some of the evidence for these economies will be summarized here.

3.1 ECONOMIES OF SCALE

The principle of economies of scale, that the per-unit cost of performing work decreases as the volume of work performed increases, appears to hold true for computation and communication networks as well as for other production systems.

Grosch is attributed with stating that the per-unit cost of information processing in computers has a square root relation to increasing size. Other investigators (Knight, 1968; Solomon, 1966; Schneidewind, 1966; Roberts, 1969) have examined economies of scale under a variety of conditions for computing equipment, although economies for complex networks have not been evaluated. One of the difficulties associated with the investigation of scale is that, since the larger, more powerful computers are also those that result from technological innovations, it is difficult to separate the effects of technological advance from the economies of scale. Roberts, for instance, indicates that central processing units enjoy a cost advantage only just greater than linear with increasing size (i.e., $P = C^{1.1}$), whereas cost performance over time doubles every few years.

Another difficulty arises in that there is a differential rate of advantage for each type of computation and communication device and each combination of such devices. The evidence for central processing units, storage devices, communication channels, and input/output devices is covered below.

3.1.1 Central Processing Units

Although technological innovations may overshadow those of mere size, some economic advantage apparently does exist in the tremendous speed of modern computers. For a teleprocessing network, the question to be asked is whether enough advantage can be gained in using a more powerful central processing unit to offset the communication channel costs of gathering together sufficient data processing work to attain and maintain an efficient utilization rate.

Practical questions concerning the advantages of relative centralization of computing may be partially answered by simulating the characteristics of actual processors in an analytic or discrete simulation model. Some insight may be gained by considering the formula

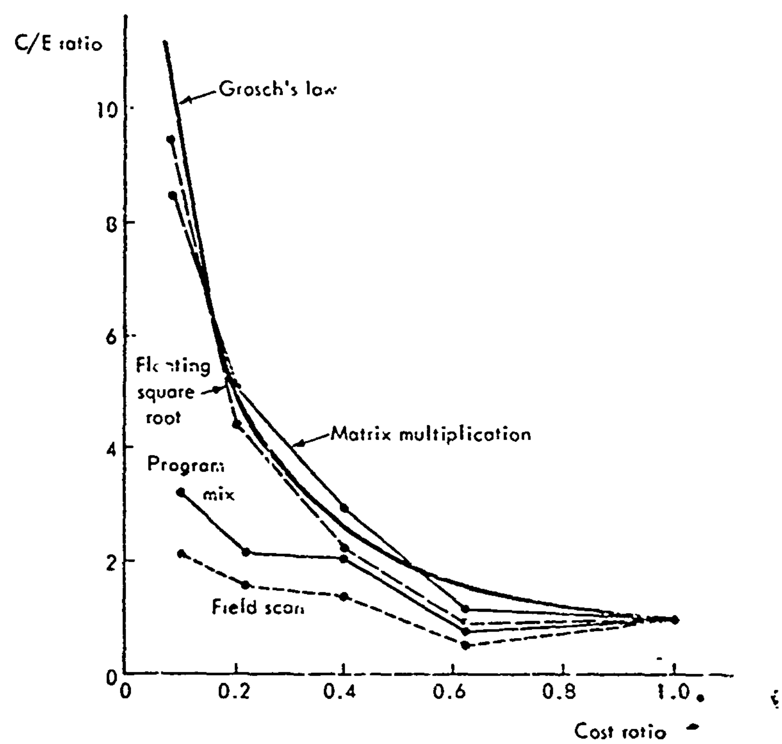
$$C/E = KC^{b}$$

where: C = cost, such as dollars per month

E = performance measure, such as megabits modified per second

K = a constant

Sharpe (1969) reports the relations shown in Figure 12 for various tasks processed on several 360 models.



Model	Cost ratio	C/E Matrix Mult.	C/E Fltg Sq. Root	C/E Program Mix	C/E Field Scan
75	1 000	1.000	1.000	1 000	1.000
65	.625	1.127	0.999	0.975	0.687
50	.400	2 879	2.162	2.250	1.426
40	.213	4.880	4.628	2.248	1.663
30	.100	8 595	9.597	3.238	2.143

Source Based on data in Solomon, "Economies of Scale and the IBM System/360," *Communications of the ACM* June 1966

Figure 12. Economies of Scale for Central Processing Units.

3.1.2 Communication Channels

While the economies of scale have not been so generally reported for communication lines as for central processing units, the cost per bit of information falls quite dramatically from a few bits per penny at low channel capacity to thousands and millions of bits per penny at very high line speeds (Figure 13) and can be expressed by the functional relation,

$$C/E = .011DE^{-.4}$$

where: C = cost

E = a performance measure

D = line length

3.1.3 Storage Capacity

A memory system for a computer may be composed of over a hundred units, each logically capable of storing a little or a lot of information. Typical storage media are: cards, holographs (including laser-based ones), magnetic cards and strips, magnetic core, magnetic discs, magnetic drums, plated wire and rods, printed pages and microfilm, punched paper tape, and thin film. Processors are normally able to communicate with only a limited number of devices, requiring other processors (such as optical character readers that pick up information stored on paper or microfilm) to transfer information from component to component within an overall memory system. Arbitrary decisions may be made concerning whether a particular component is considered to be an input or a storage device; with an appropriate tape-handling system to catalog and index the files, an entire tape library of hundreds or thousands of tapes may be considered a direct extension of the computer system memory, albeit with a somewhat lengthy access time.

Performance of memories depends on two factors: the volume of information stored and the time to access an item of information. Access time itself may be divided also into seek time--the time required to find or reach the location in a store or block of information--and the time required to read from or

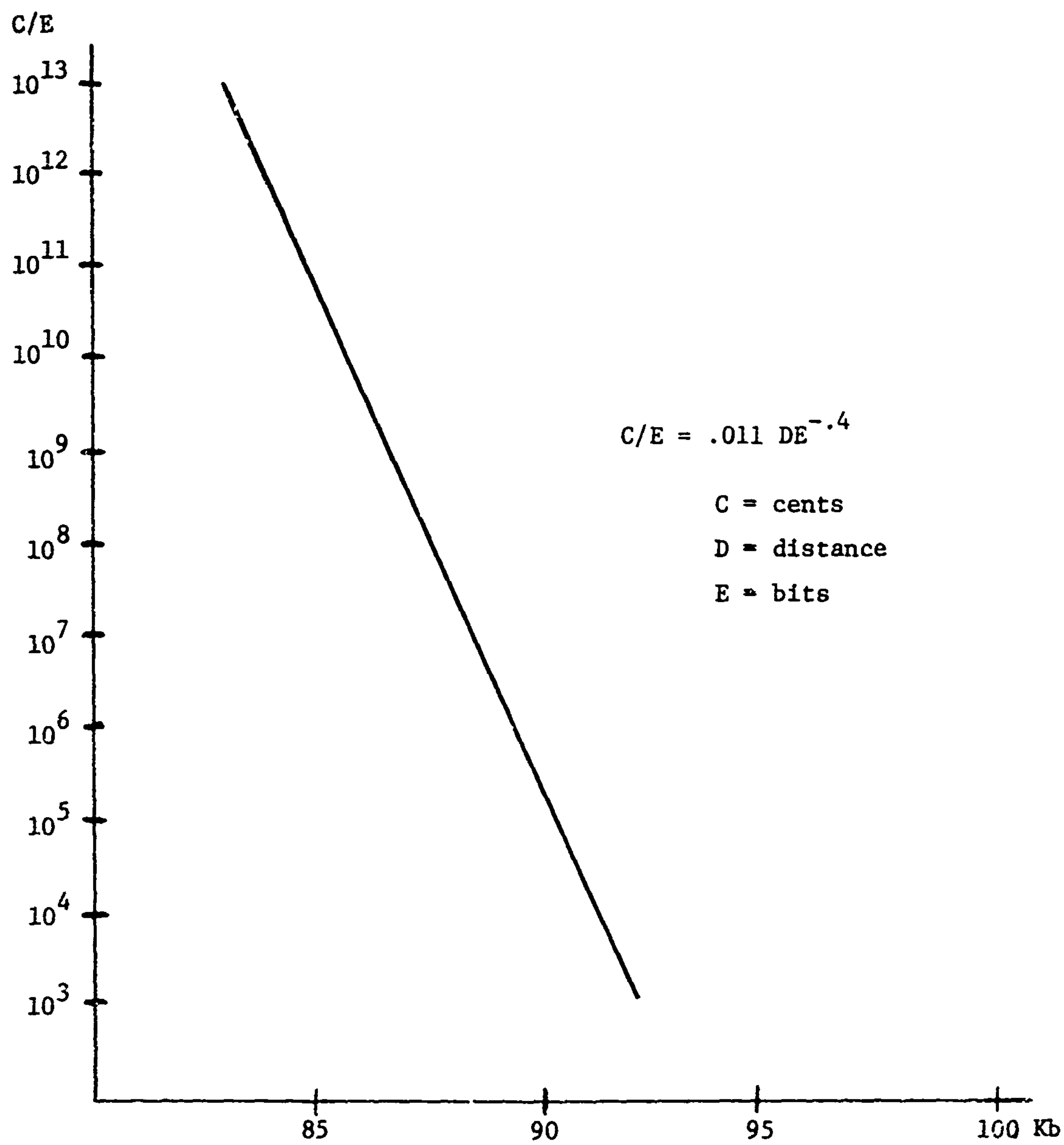


Figure 13. Economies of Scale for Communication Channels

write into the store. Storage media are usually classified as random, serial, or mixed access. For random-access devices (magnetic cores, wires, and rods), all locations are equally accessible and seek time disappears as a factor. For strictly sequential-access devices (magnetic tape, for instance), seek time is a function of the distance that the desired item lies from the starting place on the medium. For mixed access (mostly rotating devices like drums and discs, but also tape strips and magnetic cards), the time depends upon the number of read/write heads as well as rotational speeds and the physical file size.

The economies of scale for core storage vary with the size and speed (transfer rate) of the device (Figure 14). The function expressing the relation that has been found is:

$$C/E = .3\ S^{-.6}T^{-.25}$$

$$\text{or } \ln (C/E) = .3 - .16 \ln S - .25 \ln T$$

C = monthly rental

E = performance

S = bits of memory

T = cycle time

Economies of scale for discs and drums involve rotating time, total capacity (number of bands x number of bits per band), and an interlacing factor. Capacity may be added by adding more discs or drums to a single controller to obtain some cost-effectiveness gain or by increasing the total capacity by denser packing or other means. The random-access time (rotation rate) has not been greatly improved over time, but sequential read/write time (the interlacing factor) has. Costs per bit have declined significantly. Regression analysis has produced the following equation (Sharpe, 1969 p. 402).

$$\ln(C/E) = 7.05 + .17A - .66 \ln E_R - .09 \ln E_S - .5 \ln K$$

C = cost

E = bits of memory

A = years since first delivery (1967 = 0)

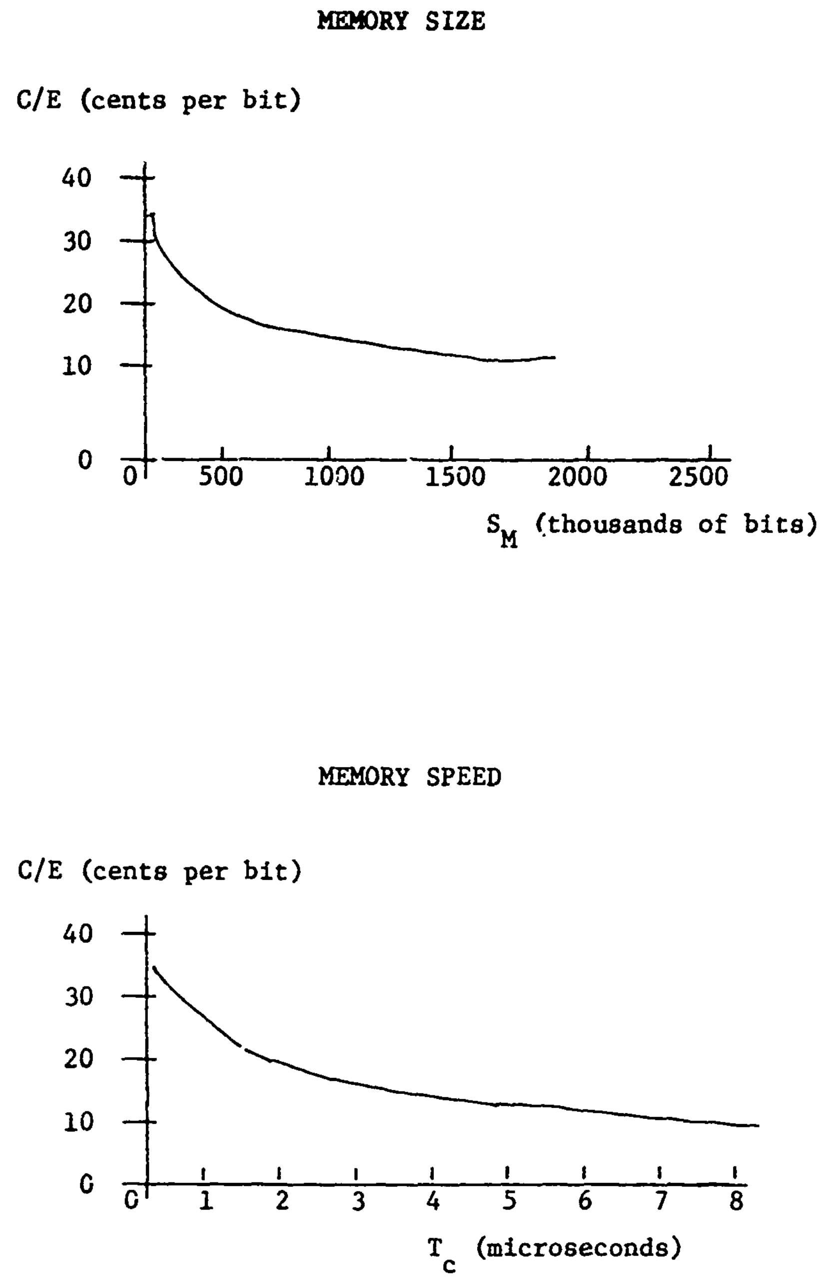


Figure 14. Economies of Scale for Core Memory by Size and Speed.

E_R = expected value of random access (seek) time, in μsecs

E_S = expected value of sequential read/write time, in μsecs

K = total storage capacity in millions of bits (bands x bits per band x no. of devices)

In essence, halving seek time raises costs per bit over 50 percent, halving sequential read/write time raises costs less than 10 percent, and doubling total capacity results in a 30 percent reduction of costs per bit. Since many means of increasing capacity and cost-effectiveness exist, there is a large variation in the cost of such storage.

To get the random-access time of a rotating device with a movable head (usually a disc), estimate of positioning time for the head must be added to the expected average rotational time. However, known regression analyses do not yield significant coefficients for seek and read times. Moderate economies of scale appear for increases of capacity ($C/E = K^{-.37}$) such that cost per bit falls by about 23 percent when capacity is doubled.

Magnetic strips, cards, and cartridges also appear to offer some economies of scale, but such devices are mechanically complex and the economies difficult to assess. Economies of scale for magnetic strip devices are plotted in Figure 15.

Magnetic tape memories show similar characteristics with the addition of multiple channels to increase the input/output speed. The regression equation for cost given by Sharpe (p. 434) is:

$$\ell n(C) = a + .66\ell n(K) + .69\ell nT + 1.50P$$

C = cost

K = total memory, in characters

T = transfer rate, in thousands of characters per second

P = proportion of potential concurrence of operations

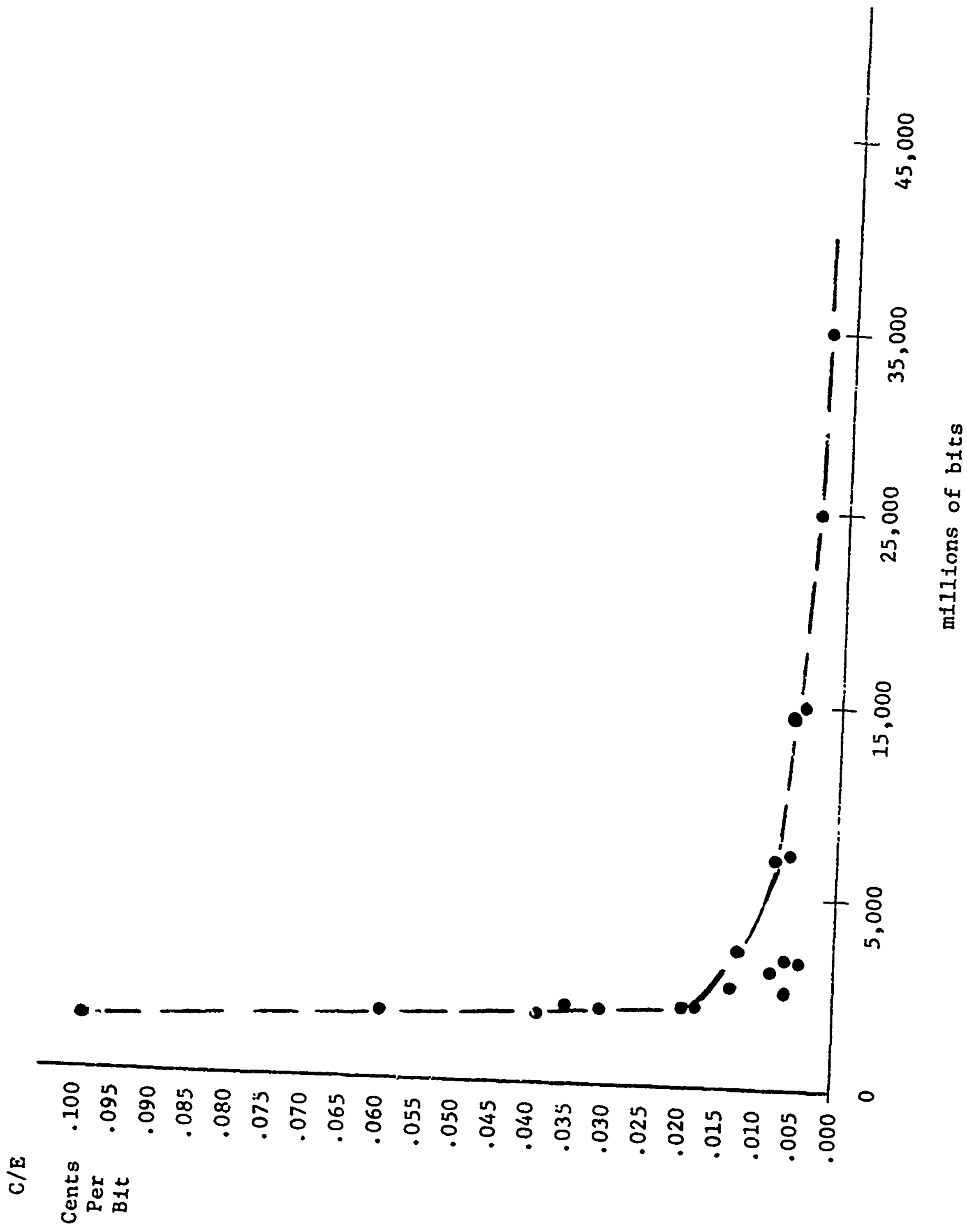


Figure 15. Economies of Scale for Magnetic Strip Devices

Doubling capacity or transfer rate increases cost by only half (60 percent), and concurrency of operation seemingly provides a great deal more economies of scale.

3.1.4 Input/Output Capacity

The cost-effectiveness of different-capacity I/O devices has not been systematically investigated. Since part of the performance measurement would be dependent upon user interface needs (e.g., faster keyboards could not be used by humans even if they were available), perhaps a definitive study may not be required. However, faster input and output impinge upon intermediate storage devices which buffer between slow I/O speeds and fast computers. It is clear that rapid optical character readers and facsimile devices operate at a more cost-effective rate (characters per dollar) than typewriters or line printers and are useful in applications requiring voluminous data. It is also true that a greater volume and range of types of data can be presented on a CRT than is practicable in either cost or time on slower-speed devices. In short, given the appropriate demand, economies of scale do exist for input/output; e.g., roughly 2.2 characters per second per dollar for a high-speed printer versus 33 characters per second per dollar for a moderately priced CRT device with a refresh rate of about 50 cycles per second--much of which could not be used by man (too fast) or camera (too slow fading of characters). More formulation will be necessary before economies of scale for control and display devices are adequately established.

3.2 ECONOMIES OF SPECIALIZATION

For each job entering a production system, there is usually some overhead cost associated with initializing (setting up) and terminating (tearing down) the job and some bookkeeping and accounting operations unless the job is precisely the same as the last one. Also, a production system providing a general capability will require more features, with their associated costs, than will a simpler, specialized system. If a variety of functions or jobs are to

be performed, the question to be answered is whether enough advantage in additional throughput and reduced overhead can be gained with specialized processors to offset the undoubtedly greater costs of several processors. In modern time-sharing systems, overhead operations may consume over 50 percent of the CPU time, and a complex executive may require half of immediate memory for efficient operation.

Work simplification programs usually operate in a two-pronged approach--through work specialization and work generalization.

In work specialization, certain processors will be elevated to a particular task or class of tasks, reducing initiation and termination costs, and special processors, both hardware and software, may be constructed that are especially efficient in performing the dedicated task. Difficulties associated with specialized processors lie, first, in finding enough work to attain reasonable utilization rates for the processors and, second, in avoiding obsolescence due to changes in the volume and technical content of the work. That is, specialized processors could be developed for many functions for a complex job. Usually, cost effectiveness depends upon having enough work to justify the expense of building the special processor and to give it a life expectancy long enough to depreciate the cost. However, with current trends toward cheaper logic, specialized processors may be cost effective even for tasks with relatively low load and short life expectancy. Some of the lowered costs depend upon the mass production of many identical elements (economies of scale); if fewer processors are built, special processors are likely to cost more per unit of power and lose some efficiency of operation. It is now believed that with large-scale integration and molecular circuits, logic units may become "throwaways"--cheaper to build than to repair.

Work generalization consists of the standardization of work elements to make as much work as possible as similar as possible in order to reduce setup and tear-down time and the amount of special handling or special processors

required. Both production processes and product elements will be standardized and generalized. Some efficiency may be lost, but in establishing general categories of work, greater opportunity for establishing an adequate load for processors will exist. Somewhat more complex processes and processors may be needed to accommodate the differences among subprocesses and products, but those should be offset by economies of scale.

3.3 ECONOMIES OF TOPOLOGY

Considerable attention has always been given to system architecture in the creation of a system organization that will foster an efficient flow of work through the system. Architecture in this sense does include both hardware and software elements, both specialized and generalized processors and dedicated and shared facilities. With the introduction of multiprocessors, integrated circuits, minicomputers, and microprogramming, the potentially cost-effective options open to the system architect have expanded greatly.

With such potential flexibility, however, system complexity has expanded to the point that special tools may be required to discover optimal, or at least operationally superior, arrangements. Some of the difficulties involved are indicated by considering the inputs to a computer installation simulation program--SCERT (Systems and Computer Evaluation and Review Technique). Over a hundred characteristics of this system, including job mix descriptions, must be specified, and the simulation considers only a few of the problems associated with netted computers.

Possibly one of the more fruitful areas of research in information processing systems may be the development of a more precise theory of operating systems and system executives. Despite a considerable investment in control theory, much remains to be discovered and, especially, integrated into the mainstream of hardware and software design. Until such a definitive theoretical basis

exists, optimum designs and optimum operations cannot be easily obtained. While network analyzers and computer facility simulators are of considerable use in evaluating alternative system configurations, more knowledge of functional relations and interactions is required to formulate potential trade-offs.

The complexity of system performance criteria introduces some further evaluative difficulties, since introducing system redundancies and quality control capabilities (error checking and reliability measures) may obscure other advantages. Requirements for processor substitution and workload balancing and reassignment are also in conflict with the cost effectiveness of specialized processors and dedicated nodes. In short, considerable operating overhead may have to be tolerated to obtain the required flexibility and survivability.

3.4 ECONOMIES OF INTEGRATION

One of the alleged advantages of automation is that it brings together in a systematic fashion the elements of an operation that have been partitioned and dispersed into many small operations. Operational redundancies, transportation delays, and additional handling are eliminated, and work is reorganized to attain efficient processing flows and arrangements. This integrative process is the principal objective of many of the new information processing systems currently being proposed by DoD agencies. That is, tasks that have been divided on functional, areal, and time bases to fit the limitations imposed by the capabilities of existing computer memories, processors, communications channels, and human operators are to be integrated into a cohesive operation using the more extensive storage and processing capabilities of modern computation and communication techniques and equipments.

Such ambitious undertakings involve a host of problems in standardization of data formats, processing procedures, and operating interfaces. A more tightly integrated system offers not only operating efficiencies but an increase in the interdependence of operations, so that a fault in one area may affect the

performance of many other parts. Consequently, integrated systems are often plagued with developmental delays and costly overruns. Hence, precautionary measures, such as developing the system in an evolutionary fashion and constructing the system of replaceable modules, are not only valuable safeguards but allow for adaptation of the system to unexpected faults and to technological and operational changes.

Integration, as evidenced by time-sharing systems, demands considerable power in executive system software. The time required to perform executive control and monitoring functions is nominally called overhead, although a complex system cannot operate in a continuous and automatic fashion without its monitor. Getting the most cost effective balance between production and supervisory operations is one of the trade-off challenges.

3.5 ECONOMIES OF QUALITY

There is no doubt that data errors, noise, and system faults and failures are expensive, not only in terms of the reprocessing and retransmission that are required but in terms of the consequences of acting upon faulty and erroneous information. However, quality in hardware and software and quality assurance procedures in operation are also expensive, and a trade-off exists between the costs of system faults and errors and the cost of preventing them.

3.5.1 Reliability and Vulnerability

Modern computation and communication equipment is quite reliable, especially when adequate preventive maintenance provisions are followed. Software is not so fault-free as hardware, but recoveries are usually easier, and steep learning curves appear for successive releases. In practice, many failures occur across system interfaces in both hardware and software. In the future, while large-scale integration promises even more reliably performing logic, microprogramming is vulnerable to firmware faults through the flexibility and complication of such a capability.

Building redundancy into the system in terms of alternative communication routes, standby computers, and backup data files plus rapid recovery and gradual degradation procedures contributes extensively to system reliability and survivability. Fail-safe, fix, and other recovery operations are overhead items involving the bookkeeping, the monitoring, and the saving of data that would not be necessary if continuity of operations could be guaranteed by other means (perfect reliability, perfect security from damage, or interdiction of information). While unnecessary precautions should not be taken, estimating what is necessary to ensure continuity of operation is a rather difficult task and not well formulated.

3.5.2 Processing and Transmission Errors

The detection, isolation, diagnosis, and correction of error are again overhead operations whose marginal utility must be evaluated with reference to the cost of the loss of accuracy in the data. The probability that an undetected error will occur in a modern computer is between 10^{-7} and 10^{-10}; the probability that logical errors will occur in checked-out computer software is unknown but is believed to be considerably greater. Typical error rates for transmission lines run from 10^{-5} in 200-band lines to 10^{-7} in megabit lines, and, with appropriate error detection and correction procedures, undetected error rates could be extended, with considerable expense, to 10^{-14}. However, transmission lines also tend to accumulate errors as a function of the number of links.* Figure 16 plots the probability of error against the number of links for both analog lines (repeaters for refreshing data) and digital lines (regenerators). (The error trade-off between analog and PCM technology is clear if costs are comparable; with LSI circuitry, an all-PCM system is estimated to cost approximately half that of a new FDM system. However, the large number of analog lines that exists tends to militate against replacement because of changeover costs.)

*Theoretically, computing systems would also accumulate errors as a function of the number of steps, routines, or functions operated, but no verifying statistics were found.

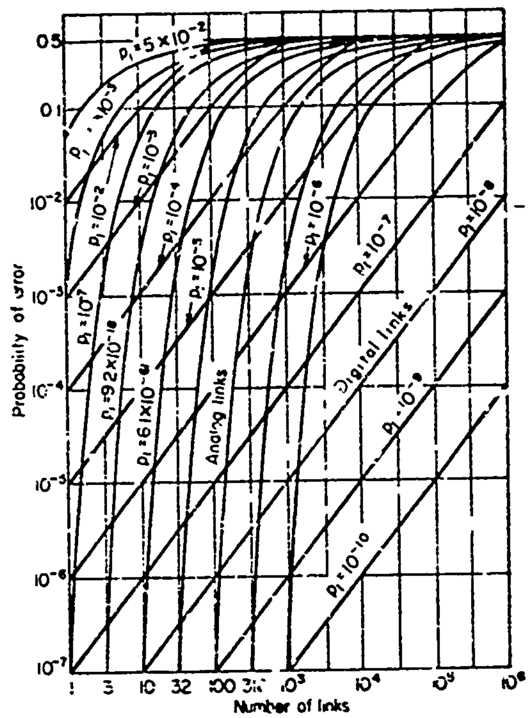


Figure 16. Probability of Error in Analog (Repeater) and Digital (Regenerator) Lines.

Most errors generated in an information system are made by humans in making inputs or interpreting outputs. While errors may be tolerated by humans in oral or printed material, including some graphic or pictorial displays, errors in numeric and other precise information cannot be. Consequently, while the value of detecting and correcting an error may vary greatly, the value of careful legality checking and of clear, unambiguous displays usually justifies the cost. It is also possible to build a tolerance for error into interactive computer programs. While such interactive programs are convenient for the human operator, they can be expensive in programming and computer time. Nevertheless, in a complex system, verification of messages through legality checking and feedback techniques is normally cost effective. Increasing the accuracy and reliability of the least effective operation in the system (i.e., that across the man-computer interface) will do more to enhance the performance of the total system than striving to improve already highly reliable and accurate processing and transmitting equipment.

3.6 ECONOMIES OF INNOVATION

The impact of technological progress on the cost-effectiveness of information processing equipment and techniques has been very great. Development of extremely complex, small, fast, and cheap logic units through LSI technology will have tremendous impact upon central processors, memories, terminal equipment, and communication lines. Similar economies in system organization and software production appear likely.

3.6.1 Central Processing Units

Much of the bit crunching in the future will be speeded by various kinds of parallelism in the computer's organization--pipelining, multiple processing elements, and associativity. More power and flexibility in the instruction repertoire of computers will be achieved by microprogramming and by moving simple machine instructions toward implementation of more powerful procedure-oriented instructions in hardware and firmware.

The cost-effectiveness of central processing units on the basis of past trends is depicted in Figure 17. An order-of-magnitude improvement in cost effectiveness has been realized approximately every six years and seems likely to continue for the foreseeable future.

3.6.2 Memories

The number and variety of devices that have been used over the years for information storage have been steadily growing and changing. The speeds, sizes, and costs of some of the more popular memory devices and those holding future promise are shown in Figure 18. The crucial development for the future seems to lie with large (trillion-bit modules), high-speed (50 nanoseconds or less), very cheap (thousands of bits per penny) laser, holographic, or bubble memories.

Although magnetic core storage is being replaced to some slight degree by such very-high-speed devices as thin-film units, it has long been closely allied with the central processing unit as part of its cost. However, speeds and costs per bit of core storage have been steadily declining with the years at an apparent rate of 15-20 percent per year. Bulk stores cost about one tenth as much per bit as do immediate memory units, but their access times are longer.

The cost-effectiveness of rotating, fixed-head auxiliary memories is depicted in Figure 19, an order of magnitude change in 8-10 years. Similar curves may be derived for memories with moving heads and for magnetic strips and cards and other semirandom access devices.

For tape drives, the relation of cost effectiveness to year of introduction is shown in Figure 20, an order of magnitude change in around 15 years. Future advances would seem to lie with parallelism in operation rather than in further exploitation of the medium, but steadily decreasing disk pack costs may displace tape units completely.

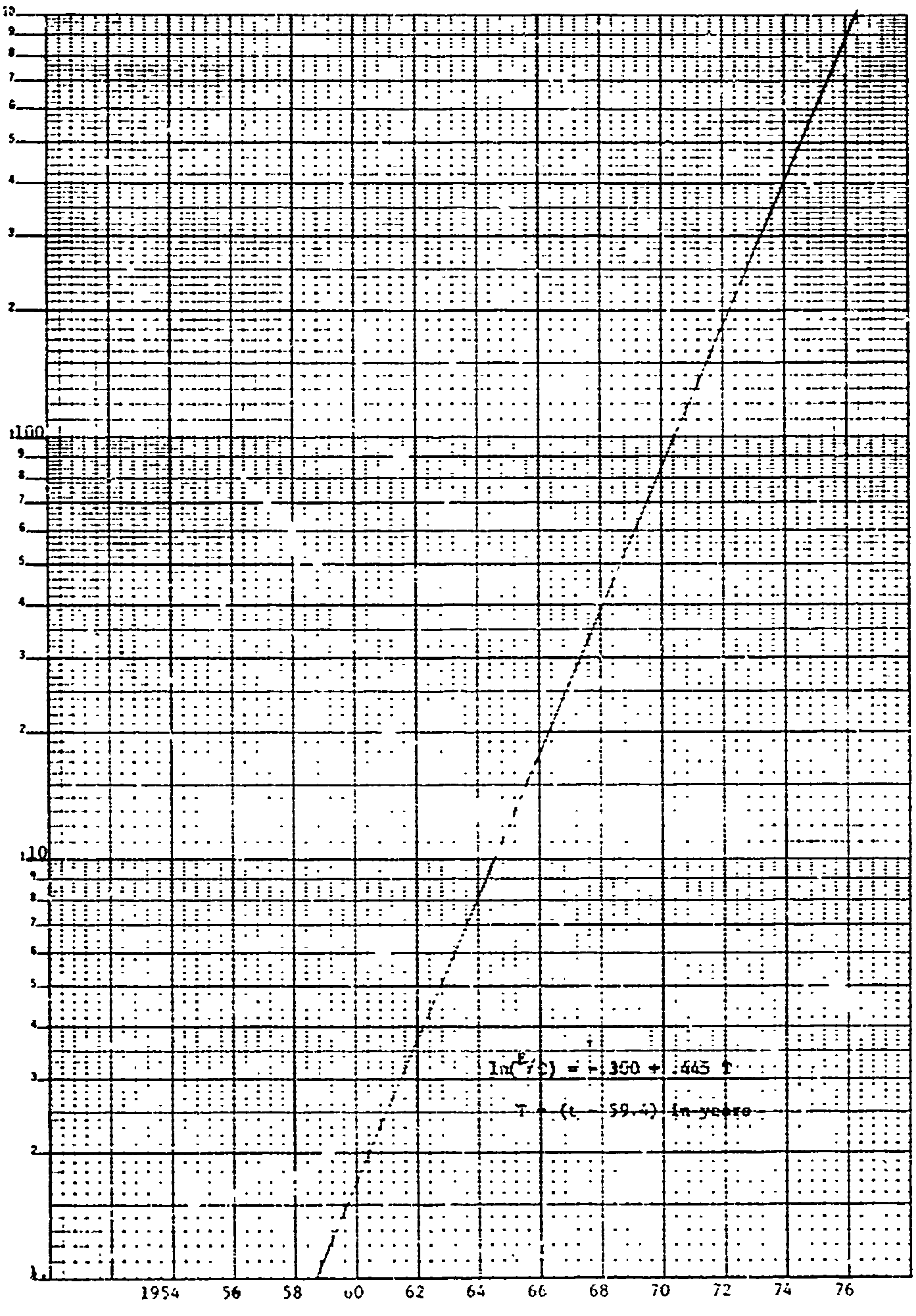


Figure 17. Increase in Cost-Effectiveness of Central Processors

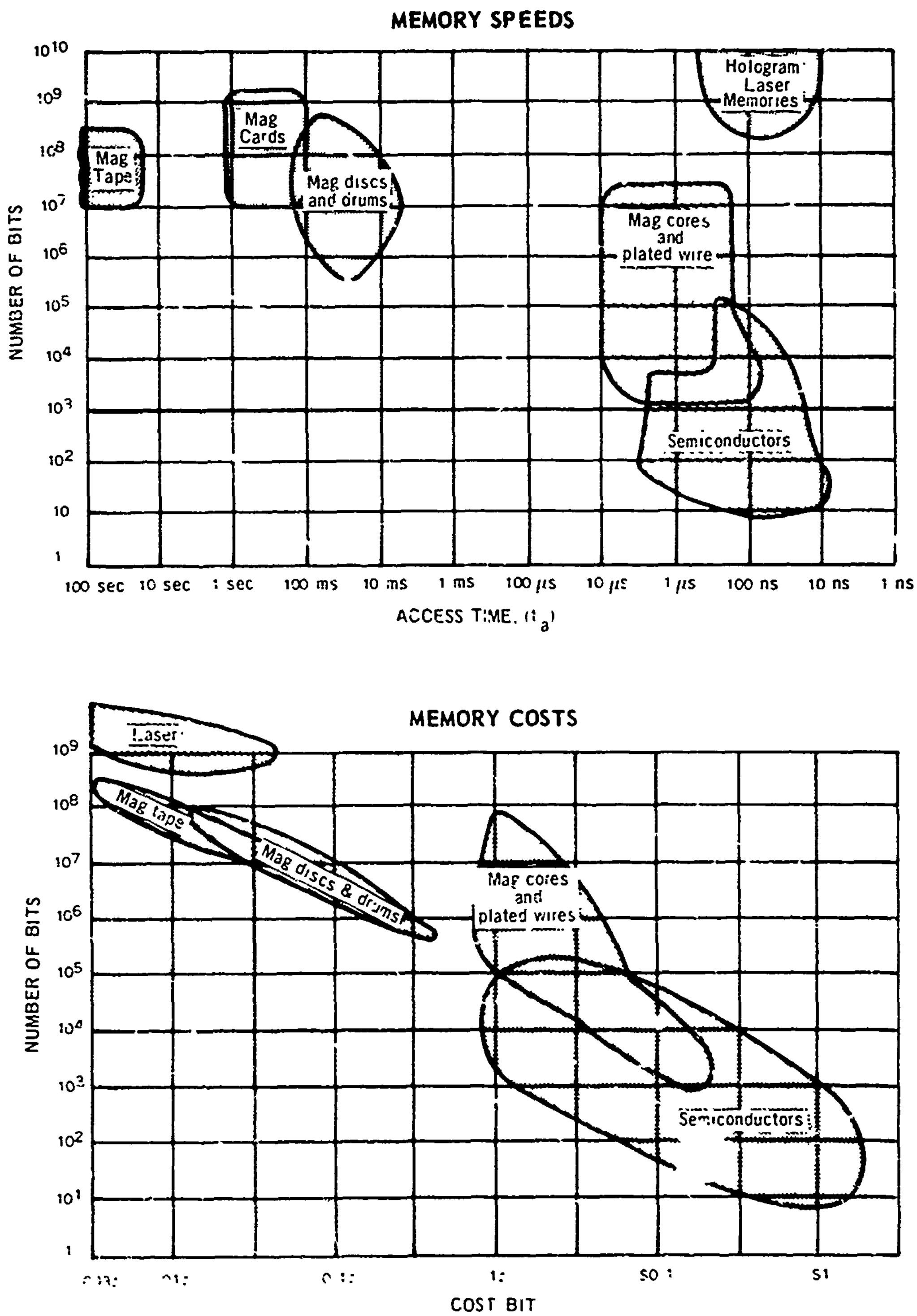


Figure 18. The Speed and Cost of Computer Storage.

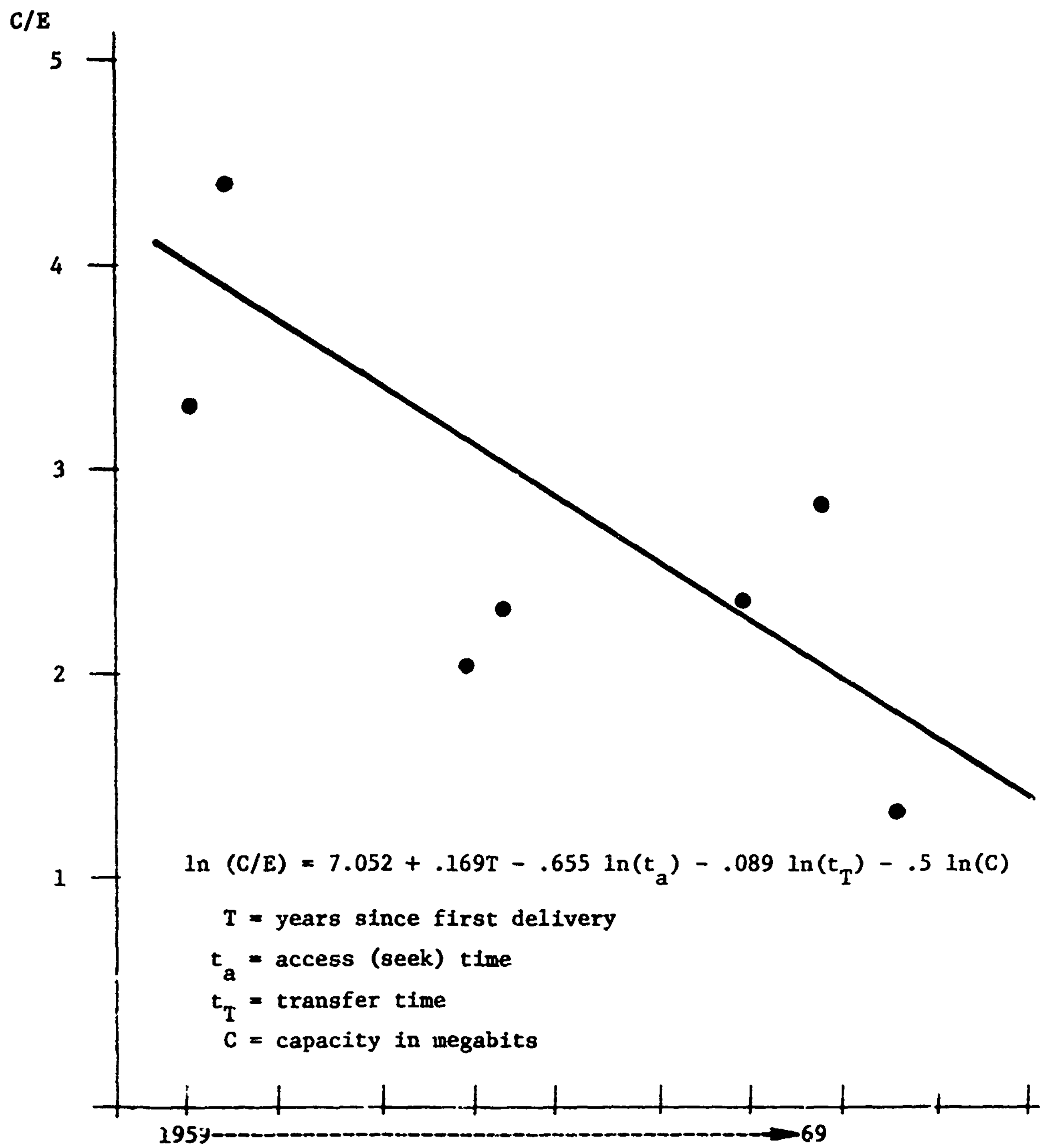


Figure 19. Increase in Cost-Effectiviteness of Auxiliary Memories (Discs)

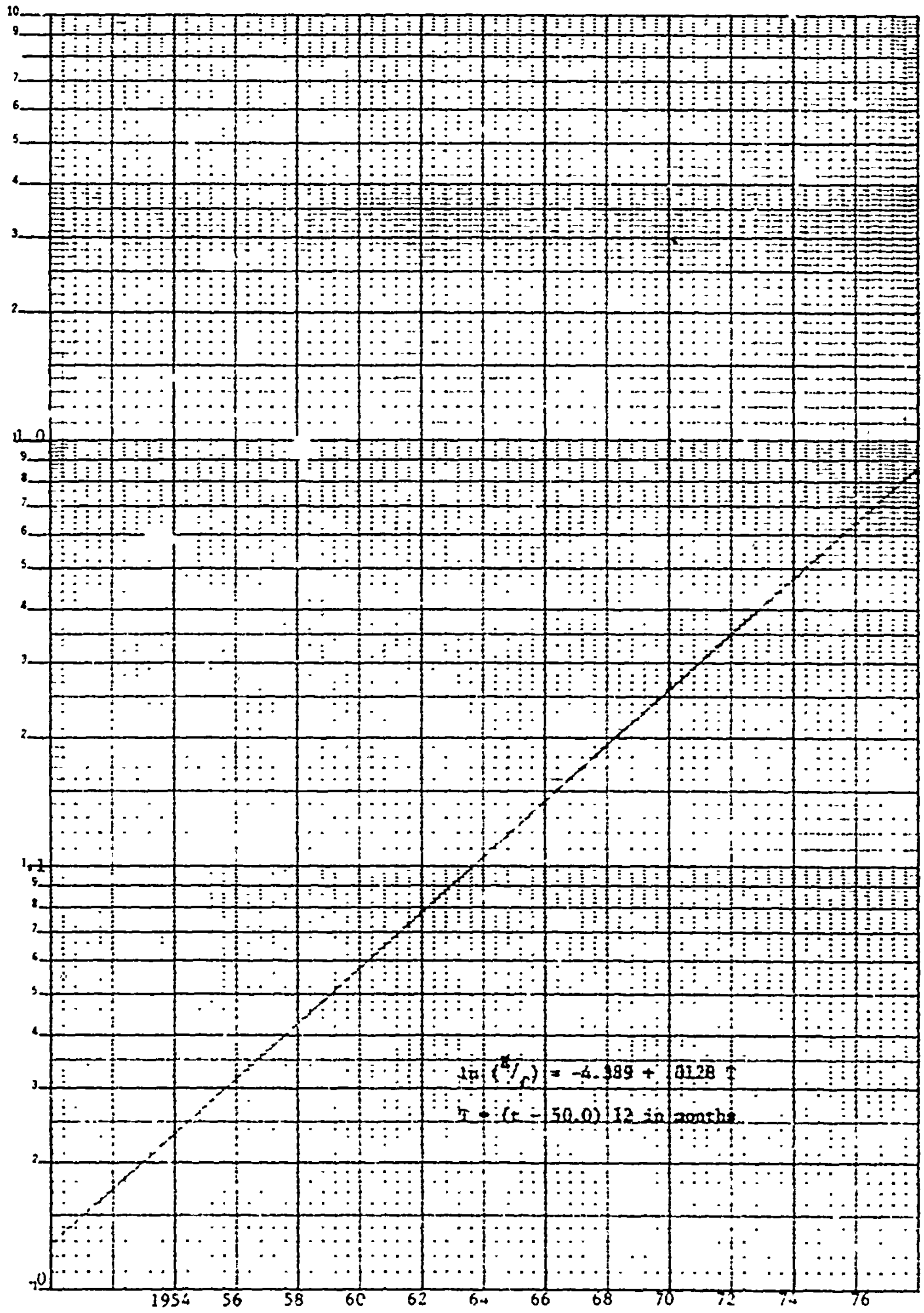


Figure 20. Increase of Cost-Effectiveness of Tape Drives

3.6.3 Terminals

Terminals present an even wider range of operating speeds than do memories. Figure 21 shows a comparision of the speeds of various peripheral devices. Terminals, however, have not been improving in cost-effectiveness as rapidly as other devices. Typewriter devices have remained almost at a standstill in cost-effectiveness although recent years have seen the introduction of some terminals with high output rates achieved by means of tiny ink jets, heat sensitivity, wire-matrix print heads, and rapidly moving character wheels.

Cost-effectiveness curves for card readers, card punches, and line printers over time are shown in Figures 22, 23, and 24, respectively. Although an order-of-magnitude gain in performance per dollar over 15 to 20 years is not trivial, it is slight in comparison to the change in less mechanical devices. The cost of CRT display devices has also fallen greatly, while their speeds, capabilities, and display capacities have grown tremendously. CRT display tubes are combined with such a variety of other interactive devices (programmable and alphanumeric keyboards, cursors, light pens, and dials) that costs are hard to evaluate.

Innovative trends have also not been assessed for graphic input devices, including CRT graphics, optical character readers, facsimile devices (including some in combination with CRT and OCR devices), microfiche, cassette tape I/O devices, and dozens of direct sensors such as radar, infrared, pressure sensitive, heat sensitive, radiation sensitive, and sound sensitive. Some of these devices are too new to have accumulated a history upon which to establish trends, and others are too specialized to be of general interest.

In the future, optical scanners, speech-to-digital converters, and high-speed facsimile devices may be expected to become commonplace. The inventory of special sensors and relatively simple devices for source data automation

SPEED RA[illegible] (CHARACTERS/[illegible]ND)	DATA TERMINAL CATEGORIES	
	INPUT	OUTPUT
10 - 100	Keyboard	Teletype
100 - 1,000	Card, Paper Tape Reader	Low-Speed Line Printer
1,000 - 10,000	Optical Character Reader	High-Speed Line Printer
10,000 - 1,000,000	Magnetic Tape	Computer-Output Microfilm

Figure 21. Operating Speed Ranges of Major Types of Data Terminals.

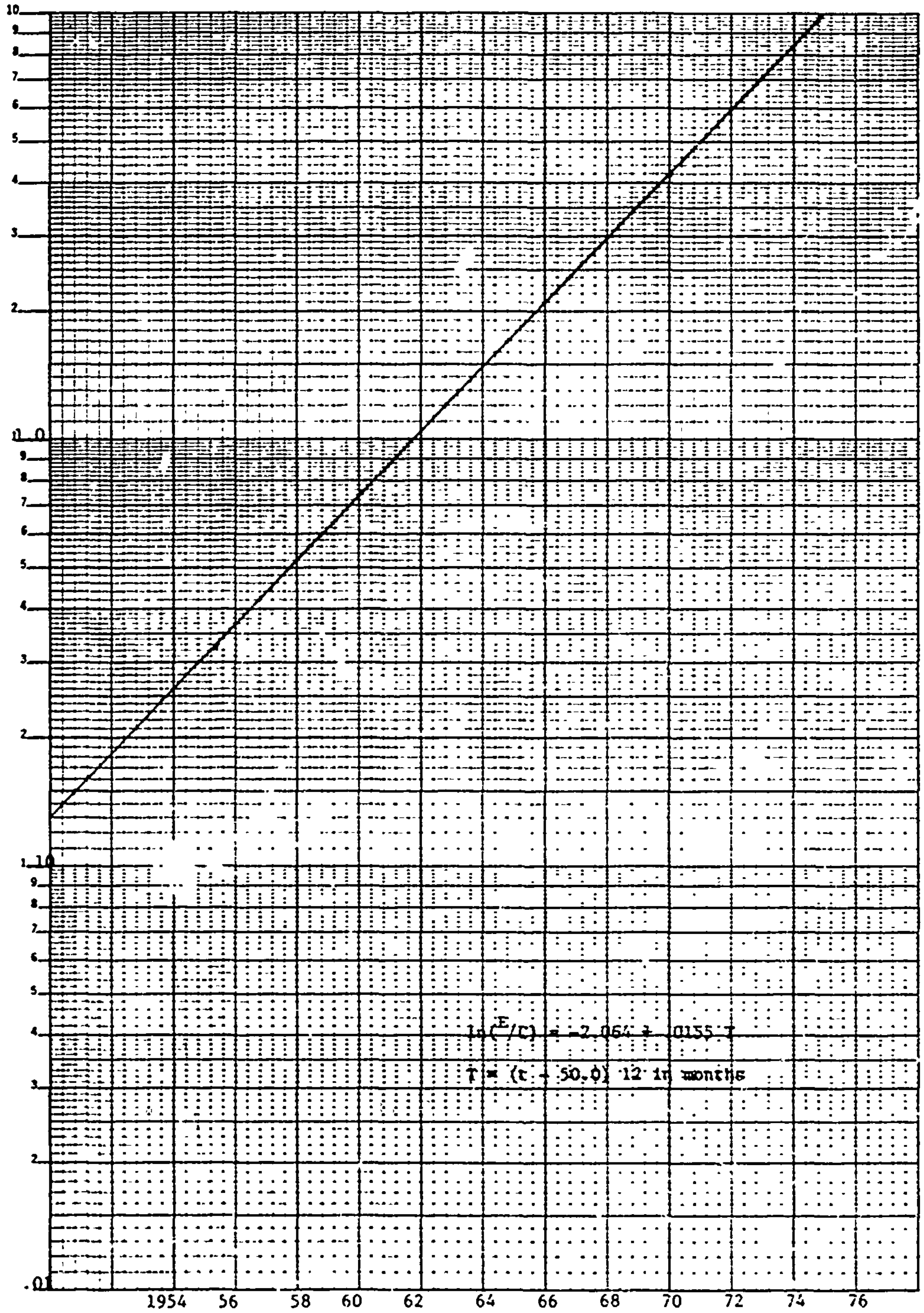


Figure 22. Increase in Cost-Effectiveness of Card Readers

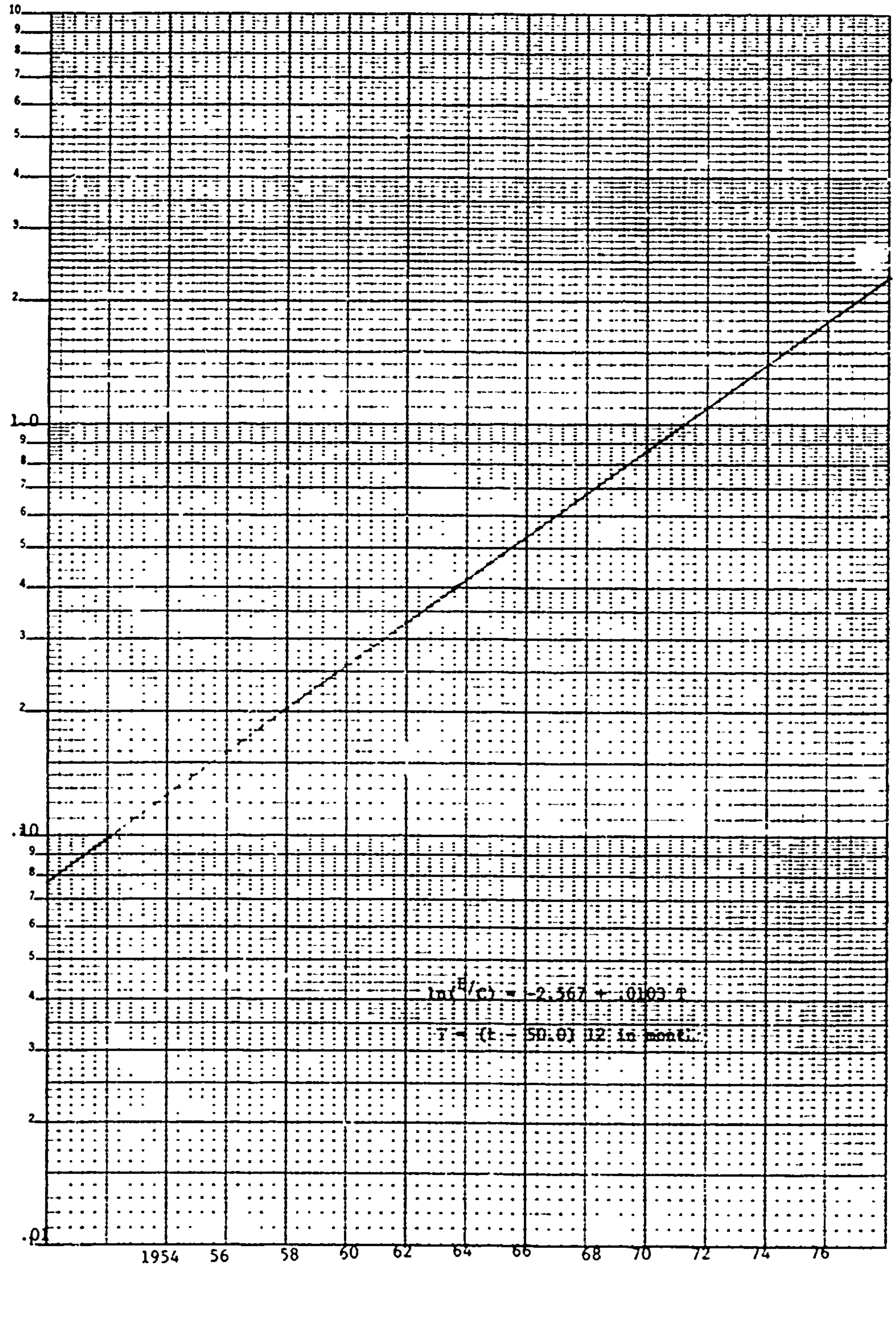


Figure 23. Increase in Cost-Effectiveness of Card Punches

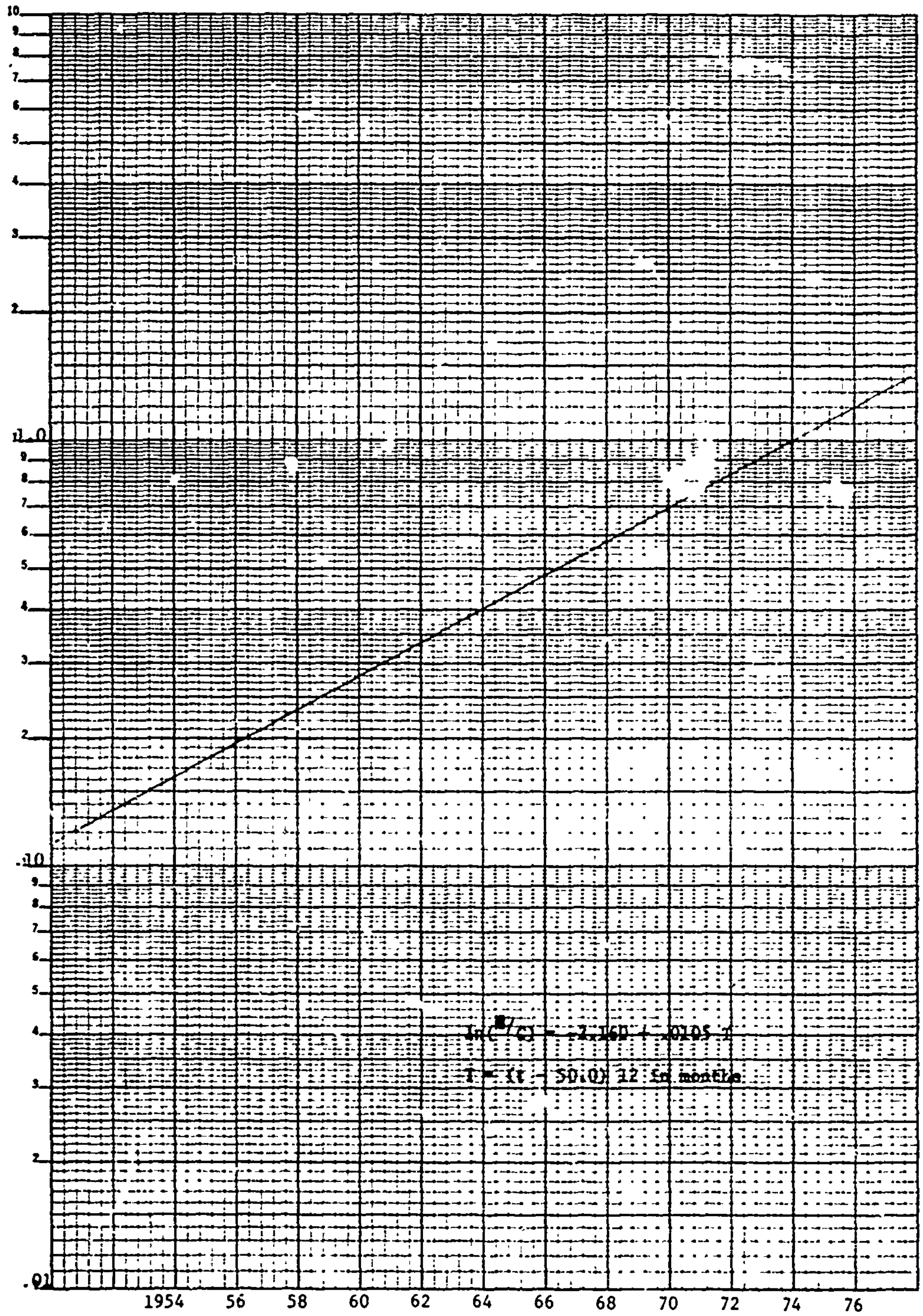


Figure 24. Increase in Cost-Effectiveness of Line Printers

(e.g., credit and other identification cards combined with simple and complex keyboard devices) is growing rapidly. The power and economy of minicomputers, combined with this plethora of control and display devices, may be the trend of the future, especially as these are combined with PCM and TDM modems for data transmission.

3.6.4 Communication Lines

The trend of technological progress in data communications has been seen in Figure 2. While the economies of scale are evident, the trends for innovation are somewhat obscured by public utility rate negotations, and a predictive formula is not yet available.

A communication system may perform many different functions in a variety of ways. Where logic is involved (as for concentration, modulating, multiplexing, switching, encryption, conversion, and formatting), prior statements concerning the cost-effectiveness of logic elements apply. A variety of such devices exist, each with special features. In installing a particular transmission system, many system-specific devices may be built. For data systems, minicomputers (or minicomputer-like devices) are being installed to perform multiplexing, formatting, and switching duties. It is expected that this distribution of logic throughout communication systems will continue.

In the future, with the availability of helical wave guides and laser devices to carry the traffic generated by videophone, facsimile, and computer interactions, lines of from 1.5- to 6-megabit capacity may commonly be installed for local loop service at a cost of about .007 cents per bit of capacity per month. Even with the high data requirements of videophone and facsimile, a tremendous amount of information must be generated to reach economical utilization rates for such lines, and a growth in multidrop and loop lines controlled by minicomputers and using time-division multiplexing must be expected. Such devices already exist, although few or no integrated systems have been constructed.

The public utilities are moving into multimegabit PCM lines for long-haul and interstation trunks as rapidly as existing FDM equipment can be economically replaced. Such lines have an extremely high accuracy rate, tremendous flexibility, and low production, operation, and maintenance costs. There are advantages in PCM for both long- and short-haul transmission. The great difficulty lies in replacing existing FDM systems with the more cost-effective PCM system. In this sense, the installation of PCM-driven videophone service promises to be of tremendous advantage to digital data processing and transmission systems. A comparison of these systems is given in Table 3.

3.6.5 Software

The cost of software has been a steadily growing proportion of all information processing system costs over the years (Figure 25). It is estimated to be 80 percent of data processing costs, the remaining 20 percent being rather evenly divided between equipment rental and salaries.

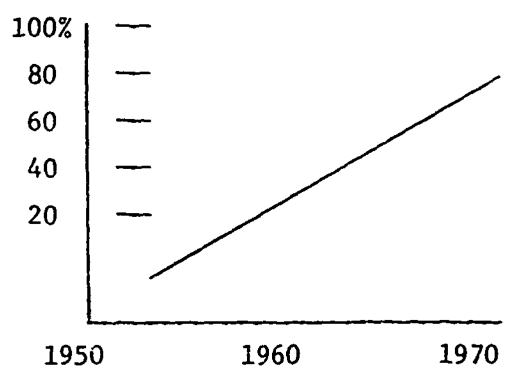


Figure 25. Proportion of Cost of Programming to Total Data Processing Costs.

TABLE 3

COMPARISON OF PCM AND FDM SYSTEMS

PCM/FM	FDM/FM
NOISE PERFORMANCE	
Noise independent of line length.	Noise accumulates with line length.
Noise independent of received signal.	Noise dependent on received signal.
Performance limited by modem design (noncritical) and code selection.	Performance variable dependent on intermodulation and noise levels, requires high quality components and rigid design.
Insensitive to interference.	Sensitive to interference.
"Sudden death" noise performance.	"Graceful degradation" up to threshold.
VOICE QUALITY	
Quantizing noise and distortion dependent upon voice level.	Noise dependent on RF signal level and intermodulation products.
Dynamic range limited by choice of basic code.	Large dynamic range limited by amplifier distortion.
Identical performance on all channels modulated (TDM).	Variations among channels modulated (FDM).
DATA TRAFFIC	
Direct digital access to pulse stream permits efficient use of capacity.	D/A and A/D conversions limit efficient use of capacity.
Supervising and signalling use same techniques as data traffic.	Special signals required for dial-up and other control signals.
SYSTEM CAPACITY	
No reduction in system capacity for data traffic.	As data traffic volume increases, power and bandwidth requirements reduce capacity.
Full modulation capabilities allocated to each time slot.	Modulation capabilities are shared in proportion to relative levels and frequencies of channels.
BANDWIDTH AND POWER	
Low power/large bandwidth.	High power/small bandwidth.
20db carrier/noise required.	40db carrier/noise required.
CIRCUITRY	
On-off, simple.	Analog, critical.
CHANNEL ACCESS	
Unrestricted channel access at all nodes.	Demodulating repeaters required.
FLEXIBILITY	
Traffic configurations set up with simple connectors and on-off circuits.	Discrete filters, amplifiers, and translators required.
All channel panels identical.	Individual units not interchangeable.
Bandwidth easily divided to accommodate many variable transceiver speeds.	Bandwidth divisions normally fixed.
MAINTAINABILITY	
Go/No-Go units: Simple tests.	Analog units: Sophisticated test equipment and careful measurement required.
Nonskilled, throwaway replacements.	Skilled adjustments required.
COSTS	
System uses basic modules, saving manufacturing costs.	Many discrete modems, etc., keep manufacturing costs up.
Power requirements low.	Power requirements high.
Simplicity of modular concept eliminates complex maintenance.	Complex maintenance of many different modules required.
Common parts and throwaway techniques reduce training and manpower costs.	Skilled adjustments and costly parts require highly skilled and costly manpower.

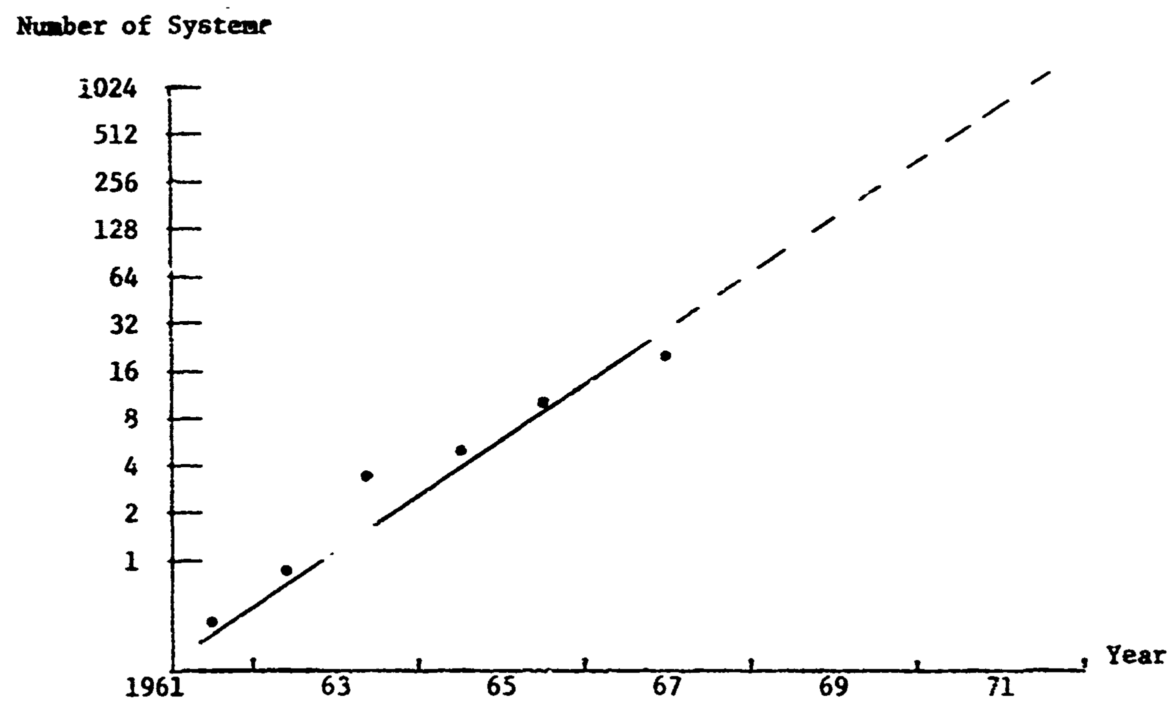


Figure 26. Estimates of the Number of General-Purpose Time-Sharing Systems in the United States.

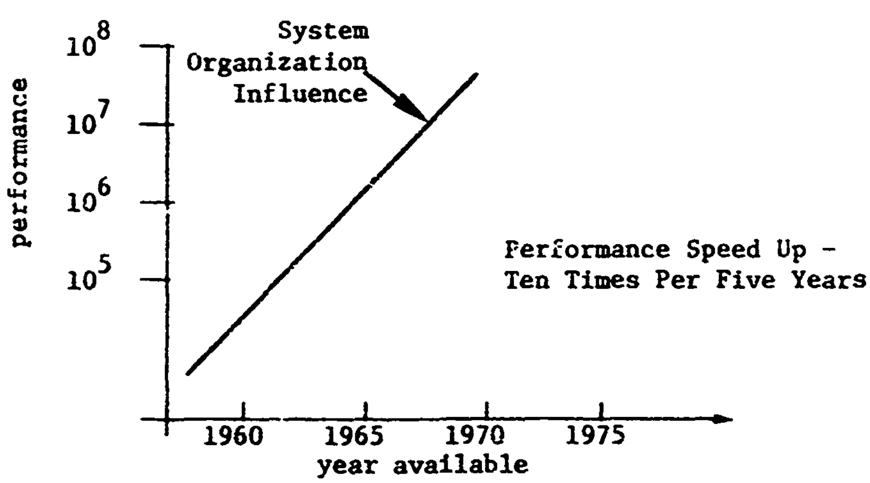


Figure 27. Computer Performance Trends (in Instructions per Second)

The operating efficiency of programs is also extremely important. The operating executive in a software system performs many supervisory and management tasks that would otherwise have to be done manually by an operator, at a large cost in idle CPU time. As the computer becomes ever larger, faster, and more expensive, more and more work must be found to keep it busy (i.e., to obtain an economic return on an investment in an expensive piece of machinery). Handling a greater variety of work and interacting with more customers results in more executive overload and input/output conversions. In some modern time-sharing systems, these overload operations consume 50-60 percent of the CPU time. While some features of current and projected hardware are designed to reduce the operating costs of executive functions, these will not entirely replace the need for more powerful and efficient operating system software techniques. The trend toward public utility or general-purpose, time-sharing systems may continue to grow and may include remote job entry service as well as interactive service (Figure 26).

The concern over executive system overhead should not be permitted to obscure the very real efficiencies in system performance introduced by the improved system organization largely achieved through system executives. Figure 27 shows that while the speed of logic has been changing by a factor of 10 every ten years, system performance has been improving by a factor of 10 every five years. Multiprogramming, multiprocessing, microprogramming, and other software capabilities may be expected to continue to influence throughput trends for some time.

Much effort is spent on writing compilers and assemblers for programming languages and other programming tools. There is little doubt that such tools increase the productivity of programmers and reduce the time and cost of producing programs, as shown in Table 4 for procedure-oriented languages vs. machine-oriented languages. Many other tools such as decision tables, program listings, and debugging tools, subroutine libraries, and other general

utility programs have similar effects. Future trends would seem to emphasize more standardization of programming languages, software "system engineering" (constructing program systems of standard, "off-the-shelf" modules), and compiler-compilers (special programs for compiling generators of programming languages). The latter development may lead to the easy generation of many special-purpose language compilers producing compatible programs that will run faster and more efficiently for producing programs for specific data processing tasks than the large, general-purpose compilers that are the mode today.

TABLE 4

PRODUCTION COSTS PER 1000 MACHINE INSTRUCTIONS FOR PROGRAMS WRITTEN IN MACHINE- AND PROCEDURE-ORIENTED LANGUAGES

	Maximum	Minimum	Std. Dev.	Median	Mean
Man-months					
123 MOL programs	100	0.14	10.18	4.00	5.89
46 POL programs	9.49	0.07	2.61	1.16	2.13
Computer hours					
123 MOL programs	294.04	0.05	42.75	15.00	29.52
46 POL programs	52.50	0.30	13.74	2.86	9.76
Elapsed time (months)					
123 MOL programs	40.00	0.06	5.81	1.33	3.55
46 POL programs	18.43	0.06	3.71	0.92	2.30

4. COMPUTATION AND COMMUNICATION TRADE-OFF ANALYSIS

Within the broad scope of the project, but limited by the rather simple prototype network analysis program that has been initially produced, a series of preliminary computation and communication trade-off analyses were performed. These analyses had the objectives of:

- Validating the network analysis model against the operations of a real system
- Evaluating technological alternatives
- Investigating the impact of computer throughput parameters

To validate the operation of the analysis program, an existing system of netted computers (the Marine Manpower Management System) was chosen. An attempt was made to recreate the real behavior of this system by modeling existing computation and communication traffic. The system response was indeed similar to the behavior of the actual system. As a practical exercise, several questions of interest to the Marines and the Project were asked of the model. It is expected that this cooperative interaction will be continued for future studies and that evaluations of real systems will be expanded to a broader sample of DoD command and control systems.

A limited sample of the possible technological alternatives was covered in initial model runs, and only a few network performance characteristics, such as network cost and response time were considered. However, the runs that were performed lend insight into the interrelation of response time and network costs for different network topologies and economies of scale. More penetrating analyses of important parameters contributing to these and other alternatives are planned.

Since the initial simple formulation of node behavior was felt to be inadequate, investigations of computer throughput were initiated, aimed primarily at evaluating the impact of various computation job characteristics, that would pave the way toward more exacting and sophisticated simulation of node behavior under a variety of conditions.

4.1 VALIDATION OF THE MODEL

Experimentation with the Marine Manpower Management System serves two purposes: first, it helps to validate the network simulation model against the operation of a real data network; and, second, it makes an immediate practical application for the CACTOS studies in computation and communication trade-offs.

4.1.1 <u>Description of JUMPS/MMS</u>

The Marine Joint Uniform Military Pay System/Manpower Management System (JUMPS/MMS) is centered in the Marine Corps Automated Service Center (MCASC) in Kansas City, Missouri, with satellite Data Processing Installations (DPIs) at seven Marine Corps bases in the continental United States and overseas. (Initial simulation runs were made using data from eight locations, but the DPI in Danang, Vietnam, has since been phased out.)

4.1.1.1 Purpose

The Marine Manpower Management System was created to:

- Improve the timeliness and accuracy of personnel and pay accounting and reporting
- Reduce errors due to human judgment and intervention
- Improve the management of the manpower appropriation
- Eliminate overlap and duplication in reporting
- Reduce the manual effort required for recording and reporting at the unit and tactical level
- Improve responsiveness to changing laws, policies, and regulations

Manpower Management Reports summarize unit strengths and attributes to enable the field commander to manage his personnel resource effectively and economically. These reports reflect changes input through Unit Diary reporting, and the output of individual records and discrepancy reports helps ensure the quality of information.

Payroll listings, checks, and accountings of leave, earnings, deductions, and allowances are provided to field disbursing officers. The data base is then available to answer queries concerning the status of members' accounts.

The Marine Corps Finance Center at Kansas City has a consolidated and up-to-date data base from which displays can be made to answer worldwide inquiries concerning pay matters. Current and historical records of overall Marine Corps disbursements and collections are kept, and reports prescribed by numerous agencies (Treasury, OMB, GAO, etc.) may be readily produced.

Financial records and accountings and analyses of findings and disbursements are available to HQMC Fiscal Division for financial management of the Corps; and appropriate personnel records, locator files, and personnel information for the broader control of Marine Corps personnel are available to the Personnel Department and G-1 at HQMC.

4.1.1.2 Network Topology

The general topology for the Marine Manpower Management System is shown in Figure 28 and Table 5. In essence, the system operates as a star net with information feeding into and out of the Marine Corps automated service center (MCASC) Kansas City. In addition to the computerized portion of the system, information is manually gathered from and delivered to outlying Marine installations about the satellite computers. Each installation has a Farrington 3030 Optical Character Reader for converting unit diaries and data transcription forms into machine-readable entries on tape. An actual physical connection does not exist between the AUTODIN Multimedia Terminals (AMTs) and the satellite DPIs. At MCASC, the AUTODIN interface is to a partition in the main processor, but both at the satellites and MCASC, data are buffered into tape storage between transmission and processing. Tapes are manually mounted, dismounted, and transported between operations.

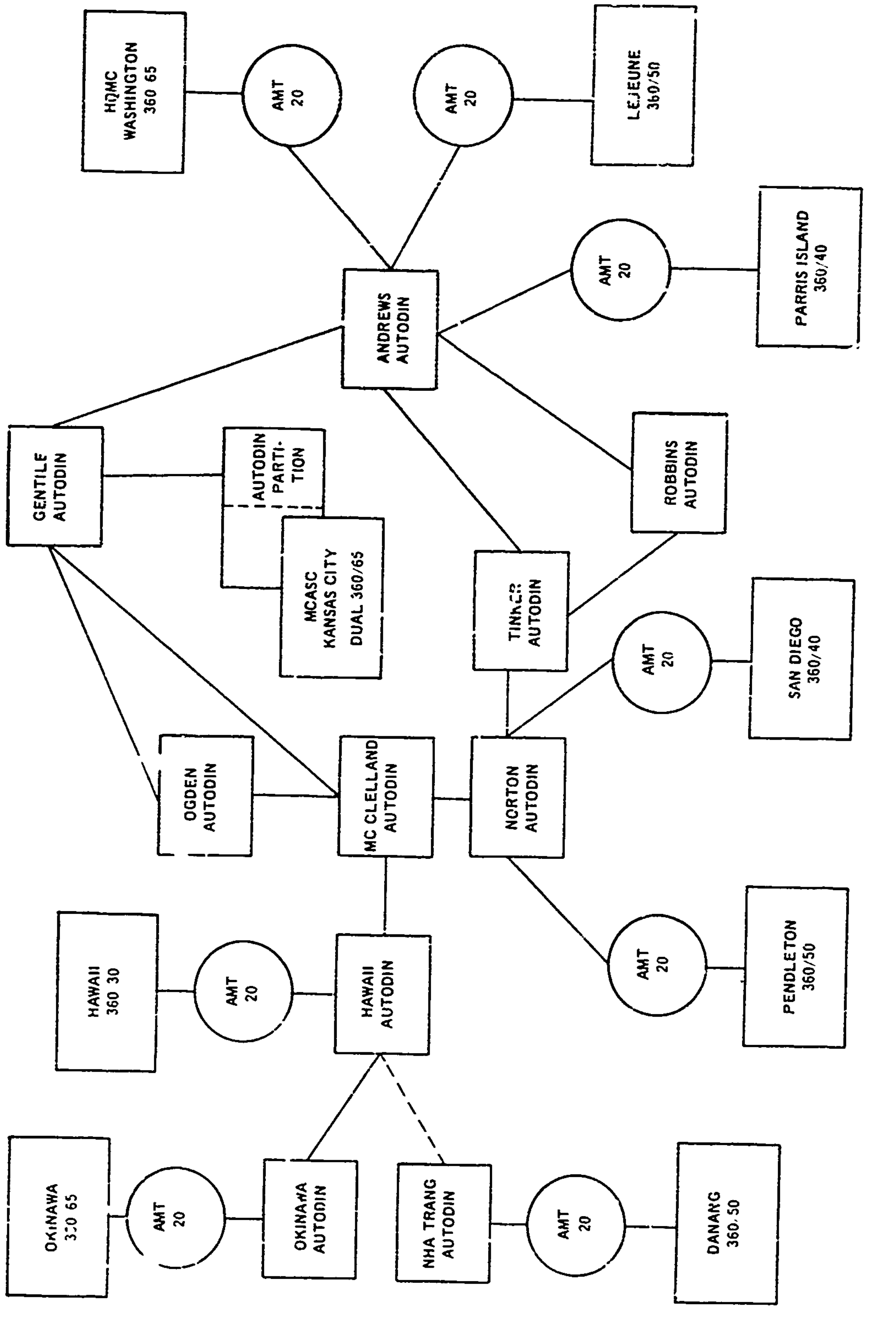


Figure 28. The Marine Manpower Management System

TABLE 5

CHARACTERISTICS OF MARINE MANPOWER MANAGEMENT SYSTEM

Location	Computer	Capacity (MBS)	Estimated Rental/Mo.
MCASC (Kansas City, Kansas)	360/65I	25.00	$50,000
CLNC (Camp LeJeune) North Carolina	360/50I	8.00	25,000
PEN (Pendleton, Calif.)	360/50I	8.00	25,000
DANANG (Viet Nam)	360/50	8.00	25,000
HAWAII (MC, FLTPAC)	360/30F	.64	8,000
OKI (Okinawa)	360/65I	25.00	50,000
HQMC (Washington, D. C.)	360/65I	25.00	50,000
PI (Parris Island)	360/40H	2.56	15,000
SD (San Diego)	360/40H	2.56	15,000
AMT (AUTODIN Multimedia terminal)	360/20/W 2701	.64	3,000

The AMTs and the AUTODIN System are connected by 1200-bps lines, except at Kansas City, whose load justifies 2400-bps lines. AUTODIN operates over cables, conditioned lines, and microwave and communication satellite channels at 2400 bps.

A consolidated data base for JUMPS/MMS is maintained at the MCASC at Kansas City. Data bases pertinent to the personnel assigned to a command or area are maintained at the satellite computers, plus additional data pertinent to the operation of the satellite. At Hawaii (Pacific Fleet), abbreviated records are kept of all Marine personnel assigned to CINCPAC. At HQMC, JUMPS/MMS data is merged with additional information to form a special data base for HQ use. Table 6 provides some hint of file sizes and activity rates for various JUMPS/MMS subsystems.

TABLE 6

JUMPS/MMS SUBSYSTEMS AND ACTIVITY

SUBSYSTEM	INPUT	MO. VOLUME	FILE	NO. ENTRIES	OUTPUT	MO. VOLUME
Bond and Allotment	Allotment Form	60,000	Allotments	122,000	Checks	511,000
			Savings Bonds	200,000	Savings Bonds	Unkn.
			Blanket Allot.	165,000	Insurance	Unkn.
			N. S. Life	24,000	Savings Plan	Unkn.
			Blanket Cos.	200	Dependent Pay	Unkn.
			B&A Accounts	900,000	Other	Unkn.
			Active B&A	511,000	Payrolls	
Retired Pay/Personnel	Posted Changes	10,000	Master Ret. File	54,000	Checks	36,000
					Payrolls	1
					Labels	
Manpower Management	Unit Diary	1,671,012 Line Blocks	Master File	233,134		2,974,000 Line Blocks
	Data Transcript					
	Class. Test Scores					
	Recruit Access.					
Reserve Pay	Changes	160,000				
Reserve Mobilization	Personnel Changes	4,000	T/O File	60,000	Orders Posted	40,000
	T/O Line Changes	500				
	Billets Changes					

4.1.1.3 Concept of Operations

Although the MMS system is operational 24 hours a day at Kansas City, the operation consists largely of ordinary batch processing for the maintenance of periodic reports. The data bases are available at MCASC for on-line information retrieval but are not available to the network. Information retrieval at both MCASC and the satellites is either via periodically produced reports or batched (off-line) retrieval runs.

Not all satellites perform identical functions. San Diego and Parris Island, as Recruit Accession Centers, generate a great deal of biodata and classification test data and produce such original records and files as service records, pay records, and medical and dental records, as well as many reports and listings of new recruits. Hawaii receives much information on personnel assigned to the Pacific Fleet area but generates only a moderate number of inputs. A duplicate MMS data base is maintained at HQMC. All record maintenance is performed at MCASC, and a copy of each record change is placed on a Touched Record Process (TRP) tape for manual transmittal (i.e., airline carrier) to HQMC. (The length of the average TRP tape--approximately 50 hours of transmission at 1200 bps--precludes direct transfer.) At HQMC, the changed records replace the old records in the data base master file. No attempt is made at HQMC to keep track of individual changes, since any one record may reflect the results of having been maintained for several changes by MCASC.

Although JUMPS/MMS is the primary application for the computers at MCASC (Kansas City), the Marine Manpower System must compete with other applications for computation and communication time at the satellites. Where the competing application is a command and control operation, the Marine Manpower System receives a lower priority. Where the local operation is deemed pressing (as, for instance, the processing of new recruits), transaction data to be transmitted to MCASC is accumulated into daily batches, or until a buffer storage unit is full (e.g., a magnetic tape), before being transmitted.

There are four subcycles in the MMS; they deal with the initiation of data at Headquarters and in the field and with the updating of data at Headquarters and at MCASC.

- Subcycle A. Data Initiation in the Field. Unit Diaries, Data Transcription Forms and other records are collected daily from all reporting centers and submitted to the satellite DPIs (SDPIs) for processing. Data Collection delay depends upon the mode and distances over which the reports must be transported. Shipborne units are normally slowest in reporting, but the number of changes is usually small. However, some reports must be carried over distances of a thousand miles or more. Unit Diaries and Data Transcription forms are typed using Farrington 3030 compatible type faces and are read onto tape by the Optical Character Reader, but some data are punched on cards for input. The tapes are edited and bad records recycled for correction. The SDPI data base is maintained on the SDPI computer and the edited records are carried from the data processing center to the communications center for transmittal to MCASC, via the AMT and AUTODIN. Pay and personnel data are normally batched and sent to MCASC during nonprime time. (Delay in transmitting a tape may range from a few hours to more than two days.) The SDPI computers are shared for all data processing applications at the stations, and MMS activities consume approximately a third of the computer time. Tapes to be transmitted must be made up into messages and line blocks (almost 400 line blocks per message and 80 characters per line block) for AUTODIN handling.
- Subcycle B. MCASC Update Cycle. At MCASC, data are received in the 360/65 partition dedicated to AUTODIN and are stored on tape for processing. Records must be stripped of AUTODIN messages and line-block control data and screened for non-MMS data. MMS records are assigned to an MMS data base update cycle and thoroughly edited before the changes are incorporated in the data base. Approximately one update cycle is run per working day (250 cycles per year); weekend time is

used to catch up if an update fails. On the basis of a two-week sample, an update consumes about 12 hours of computer time and about 27 hours of elapsed time. Changes resulting from the updates are recorded on tape and transmitted to the satellite centers for official confirmation of changes. Error lists, reports, pay rolls, and other data are also transmitted to the SDPIs, taking three to five hours in transit.

- Subcycle C. HQMC Update Cycle. The Touched Record Process tape is flown to Washington, D.C., a manual transfer taking an estimated six to eight hours. The updated records are merged into the HQMC MMS Data Base, replacing the obsolete records. A merge run averages about two hours, and approximately 20 merges are performed per month. At HQMC, MMS runs use approximately 22% of the central processor time, mostly for scheduled report generation and batched information retrieval runs and for handling Unit Diary and data transcriptions for Marine units at HQMC and at Quantico and other stations for which the HQMC MMS installation serves as the DPI. HQMC also generates numerous directives concerning personnel movements, promotions, and other changes for which it is responsible that are forwarded to MCASC for inclusion in its data files and for distribution to the affected satellites. As a Satellite DPI, HQMC processes input data basically as described in Subcycle A.
- Subcycle D. HQMC to SDPI. Changes originating at HQMC, but affecting records at SDPIs, are entered into the HQMC data base and forwarded to MCASC where they are again scheduled into a data base update cycle. From the update, tapes are made for transmission to the SDPIs and the data are forwarded via AUTODIN. Data received at a satellite communication center are accumulated on a receive tape and transported manually to the data processing center either periodically or when a full tape has been accumulated. The received messages must be sorted to separate data addressed to several systems and agencies. The MMS data are included in a satellite data base update cycle and

basic reports (e.g., payrolls and other items) printed out. At each step in this process, AUTODIN message and line block control information must be added to and stripped from the data for each transmission of the information over AUTODIN.

In summary, the total processing circuit from a satellite to the MCASC to HQMC and back takes from 5 to 10 days of elapsed time and could take more. Much of this time is consumed by manual transport and by waiting in tape storage for processing or transmission. Overall, MMS processing consumes approximately a quarter to a third of the computer time at satellites, but enough work exists to bring utilization up to a 90% load. However, the process is highly I/O bound and CPU utilization is probably much lower by an unknown amount. At MCASC, over a 30-day period, computer utilization averaged close to 75% for both sides.

4.1.1.4 MMS Traffic Pattern

Table 7 summarizes the number of AUTODIN line blocks received from and transmitted to the satellite DPIs for the period August 1970 through January 1971. Except in the case of the Recruit Accession Centers (San Diego and Parris Island), line blocks transmitted have exceeded those received. An AUTODIN line block is 80 characters in length, plus framing and check bits. Messages received by MCASC have averaged 450 line blocks per message and messages transmitted have averaged 400 line blocks.

4.1.2 Experimental Objectives and Limitations

The most important results of performing computation and communication tradeoff studies for the Marine Manpower Management System are the isolation of potential bottlenecks in the information flow and the discovery of improved topologies and/or functional allocations of tasks and traffic to improve the timeliness of system performance. Of much concern is the potential 5 to 10 day response cycle in processing a change through the system. Although data transmission and processing times are an appreciable portion of this delay, an even greater portion is due to holds or waits in the processing flow due to batching, task priorities, and quality assurance measures.

TABLE 7

NUMBER OF AUTODIN LINE BLOCKS RECEIVED AND
TRANSMITTED AT MCASC, AUGUST 1970 THROUGH JANUARY 1971

DPIs	LINE BLOCKS RECEIVED AT MCASC			LINE BLOCKS TRANSMITTED TO DPIs		
	Volume	% of Total	Monthly Average	Volume	% of Total	Monthly Average
CLNC	1,974,699	20.3	329,116	3,170,189	17.9	528,364
PEN	1,965,389	20.2	327,564	3,456,517	19.2	576,086
DANANG	1,061,492	11.9	176,915	2,468,622	13.7	411,437
HAWAII	520,308	5.3	86,718	3,754,438	21.0	625,739
OKI	995,736	10.2	165,956	1,769,619	9.9	294,936
HQMC	703,259	7.2	117,209	1,257,227	7.0	209,537
PI	1,274,682	13.1	212,447	1,201,795	6.7	200,299
SD	1,145,619	11.8	190,936	880,794	4.9	146,799

Total for the period 9,641,184 line blocks received
17,959,201 line blocks transmitted

Batch processing in and of itself is not inefficient and may be the most cost effective manner in which to maintain large files. Many personnel and payroll functions do not require very short response times although it might be most important that files be up to date when periodic reports and payrolls are produced. Holding information in temporary store until data for an efficient run is accumulated is optimally cost effective unless the delay destroys the usefulness of some of the data.

Task priorities are another matter. Without relatively complex queue disciplines to control the movement of low-priority work on a cyclic or wait-time basis, delays for low-priority work in multipriority queues are of necessity

extended. Creating a dedicated system, improving the processing speed and capacity of the existing system, or assigning higher priorities will improve the responsiveness but at some cost in funds and extra effort.

Quality assurance is of considerable importance, of course, and any effort to increase accuracy (eliminate data errors by editing and testing) usually pays for itself in reduction of reprocessing and retransmission costs. Lengthy waits, however, may result from extensive manual and machine checking of intermediate steps in the processing cycle. While all such delays cannot be avoided, more computerized assistance in error detection and correction and more automated exception procedures for faulty data can speed response times and may, over time, improve the cost effectiveness of operation. MCASC now does extensive machine checking to ensure the accuracy of sensitive pay and personnel matters, and much of the traffic in the system is caused by processing records for verification and recheck.

4.1.2.1 Potential Trade-Offs

a. Batching. Our present simple model does not permit, except through model approximations, a direct comparison of batch with time-share operations. Under the current concept of OCR source data automation, daily batching of Unit Diary and Data Transcription Form inputs seems the logical operational mode. If an interactive input capability (for example, a CRT display with keyboard and cursor associated with either one of the satellite computers or a local minicomputer) were to be delivered to field units, then data entry, editing, and verification could be done in real time. Inputs could be assigned, either manually or automatically, to processing priorities. Inputs requiring immediate action could receive it, while those permitting delayed action could be accumulated for more efficient runs, as for instance in the maintenance of large files or the production of large reports. Providing immediate access to field units does

reduce the security of the data bases against insertion of faulty information and unauthorized access. Additional checking and security measures would undoubtedly be required.

Assuming that 10%-30% of MMS transactions justify advanced priorities and that an interactive environment imposes a 20% processing penalty on the central processor, what response times may be expected for real-time versus batched tasks?

b. Priorities. Although simulation of a multipriority system is beyond our present network simulation model, the comparison of an MMS-only system (i.e., MMS has top priority for processing and transmission) versus a fully loaded (but non-priority) system provides some indication of the relative merits of a dedicated or top priority versus a shared or low priority system.

c. Redistribution of Processing. MCASC is faced by several considerable increases in its processing load as JUMPS, the advanced pay system, becomes operational. Since the computer use is already high, some increase in responsiveness might be gained by dividing the workload among several centers, either by allocation of particular functions to a center or by some other division of the work.

4.1.2.2 Model Approximations

Modeling a batch processing system with manual intervention steps presents some difficulties, since the network analysis program being used assumes an on-line interconnected network of computers operating in a continuous, interactive manner. In the network analysis, runs were made that follow these approximations:

1. A mean message size was computed from the AUTODIN traffic data for MMS and applied to all loads (114,320 bits or 256 line blocks per message).

2. To get average job and message arrival rates, the numbers of line blocks transmitted to and received from a satellite DPI on an average day were divided by the mean message size to obtain a message-arrival and a job-arrival matrix, assuming one input and one output message per job performed at a computation node.

3. Job size was computed for each processing node by calculating the number of instructions required per job to consume the amount of CPU time reported to be used, on an average, for a JUMPS/MMS data base update at the node.

4. Additional job and message traffic was included on appropriate runs to bring the total load up to the utilization rates required by the comparison.

5. Temporary storage and manual intervention holds were not simulated beyond the queueing that resulted from the load imposed on the network.

6. Local loop traffic between the computer and its peripheral devices (including OCR and tape storage as well as keyboards, card readers, and printers) was not simulated.

7. Satellite DPIs were treated as if a physical machine/machine interface existed between the computers and the AUTODIN communication system, and the AUTODIN Multimedia Terminals (IBM 360/20s and 2701 interface device) were ignored.

8. Auxillary storage requirements and access times were ignored (i.e., infinite queues were permitted and intermediate stores were ignored).

4.1.3 Dedicated Versus Public Operation

The present operation of the Marine Manpower Management System is essentially a shared network, where many other users patronize the AUTODIN digital communication system and share the computer. A basic question arises immediately: With other applications often taking higher priority, what is the effect of the shared operation on the speed and efficiency of the Marine net?

In Table 8 the speed of operation of the shared system is compared with that of a dedicated system. One column shows the computer times taken at MCASC and, on an average, throughout the network to process JUMPS/MMS data in a system dedicated to Marine personnel business only. (These values are roughly those reported as being required to process one update cycle at MCASC and at the satellites.)

The other column shows the computer times for the shared system loaded to 90% of capacity. Not much change is seen at MCASC, where the computer is already quite busy with MMS applications. However, the network average jumps beyond the MCASC response time, largely because wait time is longer for the smaller computers.

The response time for communicating is more dramatically affected, jumping from around one hour to 19--slightly beyond the actual average delay of 18 hours currently experienced. While the average delay for single messages having equal priorities is still only about half an hour, a communication delay of 18 hours for low-priority traffic could reasonably occur with a 90% utilization rate. It would appear that the response (turnaround) time is approximately one day for a dedicated system and four days for a shared one. (In either case, of course, no manual operations have been included in the model. Actual response times for all operations could be at least double those reported.)

TABLE 8

OPERATING TIMES FOR DEDICATED VS. SHARED OPERATIONS

(Hours:Minutes)

	Dedicated MMS Only	Shared 90% Load
Single Message	:11	:37
All Messages	1:10	19:25
MCASC Average		
Single Job	:31	1:33
All Jobs	20:42	23:34
Network Average		
Single Job	1:06	3:59
All Jobs	16:50	27:04
Total Response		
MCASC	21:52	42:59
Network	18:00	46:29

4.1.4 Priority Division of Work

If the cost of a dedicated system makes it infeasible, and if having fast response to some subset of functions is eminently desirable, arranging for priority processing of that subset may be the best solution. The results of arranging top-priority processing for 10, 20, and 30 percent of the MMS jobs and messages may be seen in Table 9.

TABLE 9

PRIORITY DIVISION OF JOBS

(Hours:Minutes)

	MCASC	NET AVERAGE
PRESENT LOAD		
Communication	19:25	19:25
Computation	23:34	27:04
Total Response	42:59	46:29
PRIORITY HANDLING		
10% PRIORITY JOBS		
Priority Messages	:20	:20
Priority Jobs	2:46	2:50
Priority Response	3:06	3:10
Remain Messages	19:05	19:05
Remain Jobs	20:22	21:16
Elapsed Time	42:33	43:31
20% PRIORITY JOBS		
Priority Messages	:38	:38
Priority Jobs	4:45	4:02
Priority Response	5:23	4:40
Remain Messages	18:57	18:57
Remain Jobs	17:57	18:06
Elapsed Time	42:17	41:43
30% PRIORITY JOBS		
Priority Messages	1:16	1:16
Priority Jobs	7:22	6:09
Priority Response	8:38	7:25
Remain Messages	17:27	17:27
Remain Jobs	15:40	15:40
Elapsed Time	41:45	40:32

Even assuming a 20% penalty for processing the priority tasks, enough wait time is avoided to decrease total response time somewhat. Note that when the MMS load is less than that at MCASC, the network average benefits more from priorities than does MCASC, which has a substantial MMS computation load.

This scheme would promise relatively rapid reactions in the system for priority tasks, largely by avoiding the long queues in the transmission system. However, unless rapid response for priority task does have a high benefit rating, the expense of reprogramming to save a few hours overall does not seem justifiable.

4.1.5 Distributed Processing

In view of the heavy load concentrated at MCASC and the somewhat lighter loads at the satellites, some greater efficiency might be realized by sharing some of MCASC's load with one of the satellites. This implies that there are tasks or functions that could be transferred to one or more of the satellites.

To test this hypothesis, the MMS job and message loads for MCASC and HQMC were summed and reassigned equally. The results are shown in Table 10. For the 90% utilization condition, jobs were added to reach the high utilization rates again. The results are disappointing. The message traffic shifted from MCASC to HQMC was not great enough in comparison to other traffic to make a difference. The MCASC load was lightened, somewhat, but hardly enough to justify the changes in handling.

TABLE 10

COMPARISON OF DISTRIBUTED VS.
CONCENTRATED PROCESSING
(Hours:Minutes)

	Shared Load	Present Load
90% Utilization Rate		
All Messages	19:25	19:25
All Jobs, MCASC	22:53	23:34
All Jobs, HQMC	22:53	22:53
All Jobs, Net Average	25:51	27:04
Total Response Time	45:16	46:29
MMS Only		
MCASC	18:59	20:42
HQMC	18:59	16:31
Network Average	17:07	17:08

4.2 EVALUATION OF ALTERNATIVES

Although the ultimate goal of the CACTOS project is to evaluate technological alternatives with precision and in detail, these preliminary investigations were aimed as much at gaining a better understanding of network behavior and the fundamental relations among key parameters as at evaluating alternatives. For any given information net, a unique set of requirements exists, and no one configuration of network topology, traffic patterns, and functional capabilities will satisfy all such sets of requirements in a cost-effective manner. However, although unique requirements may demand unique network configurations to obtain optimally cost-effective performance, there must be some fundamental interrelations of network characteristics that will guide the network designer toward optimal solutions. Because of the complexity of the task, a grasp of those interrelationships must exist before decisions can be made among technological alternatives; otherwise, erroneous assumptions may easily be made. Hence, exploratory investigations over the response surface are as valuable as precise evaluations of alternative network choices in a specific situation.

4.2.1 Experimental Parameters

To observe the topological, traffic, and processing characteristics, these parameters must be varied in relation to one another. In the following explorations, four choices of a six-node network; two traffic patterns, two job sizes, and one message size; and three load levels and three capacity levels were used.

4.2.1.1 Network Topology

The four network configurations, each of six nodes, are shown in Figure 29. The first three of these nets represent increasing levels of connectivity or node articulation and, hence, decreasing vulnerability. The last is the star net frequently favored for centralized computing and switching centers. Theoretically, increasing connectivity ought to reduce response time, decrease vulnerability, and increase reliability, but at additional cost. The fit of these topologies to traffic patterns, however, will probably determine cost-effectiveness.

4.2.1.2 Network Traffic

Traffic was varied both in volume and distribution. Volume was controlled by varying, from one to 25, the assumed number of on-line users at a node. Traffic among nodes was assumed to be either evenly distributed (an equal number of messages and jobs flowing between every pair of nodes) or concentrated (one centralized computer performing all teleprocessing jobs). Figure 30 depicts these conditions. Network comparisons were rendered either computation bound or communication bound by varying job and message size combinations. For the computation-bound condition, a job size of one megabit was employed; for the communication-bound condition, one tenth of a megabit was used. Message size was 400 bits for both conditions.

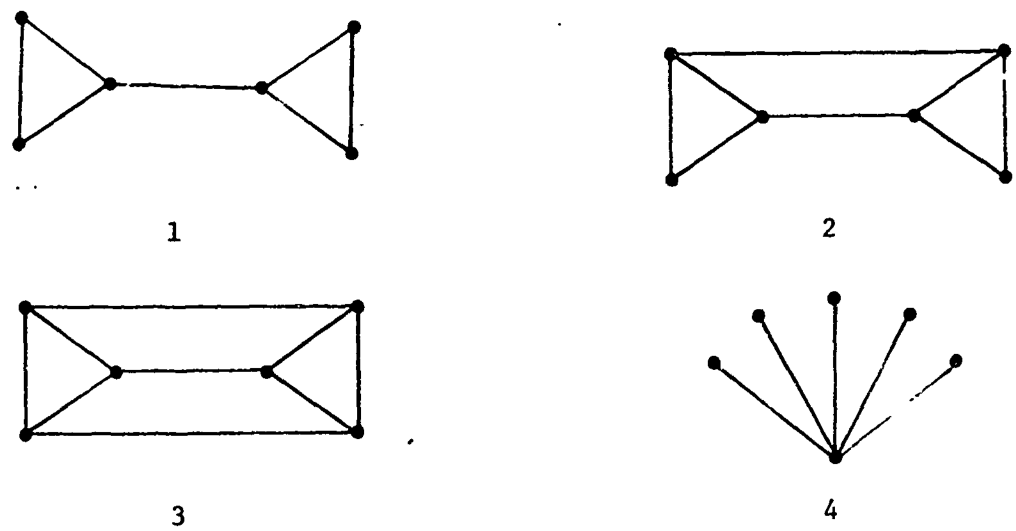


Figure 29. Experimental Network Configurations

Distributed Computation	Concentrated Computation
$k \times \begin{bmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 5 & 1 & 1 & 1 & 1 \\ 1 & 1 & 5 & 1 & 1 & 1 \\ 1 & 1 & 1 & 5 & 1 & 1 \\ 1 & 1 & 1 & 1 & 5 & 1 \\ 1 & 1 & 1 & 1 & 1 & 5 \end{bmatrix}$	$k \times \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 5 & 5 & 0 & 0 & 0 & 0 \\ 5 & 0 & 5 & 0 & 0 & 0 \\ 5 & 0 & 0 & 5 & 0 & 0 \\ 5 & 0 & 0 & 0 & 5 & 0 \\ 5 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}$

Number of users at each center using remote computers:

1 User $\rightarrow$ k = 288

5 Users $\rightarrow$ k = 1440

25 Users $\rightarrow$ k = 7200

<u>Assumptions</u>: Each remote user enters a 10-byte message and receives a 90-byte reply every 60 seconds.

Figure 30. Traffic Specifications

4.2.1.3 Computation and Communication Capacities

The capacities of three computers (IBM 360/30, 360/40, and 360/65) were specified for the nodes, and three channel capacities (2400, 4800, and 50,000 bps) were specified for the links in the network. Costs for these combinations plus the costs of appropriate switching and communication interface devices are shown in Table 11.

TABLE 11

NETWORK CONFIGURATION COSTS

Capacity Assignments*	Config. 1	Config. 2	Config. 3	Config. 4
a A1	$101,034	$102,426	$103,818	$ 98,258
b B1	101,909	103,426	104,940	98,875
c A2	172,665	184,290	195,915	149,415
a B1	145,974	147,366	148,758	143,190
a C1	290,184	291,576	292,968	287,400

*Key: Job Processing Rate	Computer	Switches
a= 2.4 Kb	A=360/30	1=Low speed
b= 4.8 Kb	B=360/40	2=High speed
c=50.0 Kb	C=360/65	

4.2.2. Network Performance Characteristics

The principal performance measures evaluated in these analyses are response times and costs for computation, communication, and the network totals. Within the framework of response time and cost, topological and traffic variables were considered.

4.2.3 Results

Results of the computer analysis of the experimental data using the CACTOS model are shown in Tables 12 and 13. Three values are given for the response time for each combination of parameters. They are: Total Response, Communication Response, and Computation Response.

TABLE 12

CACTOS EXPERIMENTS: RESPONSE TIMES FOR JOB SIZE OF .1 Mb, DISTRIBUTED TRAFFIC USING DIFFERENT CONFIGURATIONS AND LOADS

Configuration			1	1	2	2	3	3	4	4
Number of Users			1	20	1	20	1	20	1	20
JPR (Mbs)	LCC (Kbs)	RT (Secs)								
.547	2.4	Total	.498	.562	.451	.493	.428	.464	.475	.531
		Comm	.314	.354	.267	.285	.244	.256	.291	.322
		Comp	.184	.208	.184	.208	.184	.208	.184	.208
.547	4.8	Total	.347	.380	.323	.351	.311	.338	.335	.366
		Comm	.163	.172	.139	.143	.127	.130	.151	.158
		Comp	.184	.208	.184	.208	.184	.208	.184	.208
.547	50.0	Total	.211	.235	.207	.231	.205	.229	.209	.233
		Comm	.027	.027	.023	.023	.021	.021	.025	.025
		Comp	.184	.208	.184	.208	.184	.208	.184	.208
1.602	2.4	Total	.377	.419	.330	.350	.306	.321	.353	.388
		Comm	.314	.354	.267	.285	.244	.256	.291	.322
		Comp	.063	.065	.063	.065	.063	.065	.063	.065
1.602	4.8	Total	.225	.237	.201	.208	.189	.195	.213	.223
		Comm	.163	.172	.139	.143	.127	.130	.151	.158
		Comp	.063	.065	.063	.065	.063	.065	.063	.065
1.602	50.0	Total	.089	.092	.086	.088	.084	.086	.088	.090
		Comm	.027	.027	.023	.023	.021	.021	.025	.025
		Comp	.062	.065	.063	.065	.063	.065	.063	.065
25.9	2.4	Total	.318	.358	.271	.289	.248	.260	.295	.326
		Comm	.314	.354	.267	.285	.244	.256	.291	.322
		Comp	.004	.004	.004	.004	.004	.004	.004	.004
25.9	4.8	Total	.167	.176	.143	.147	.131	.134	.155	.162
		Comm	.163	.172	.139	.143	.127	.130	.151	.158
		Comp	.004	.004	.004	.004	.004	.004	.004	.004
25.9	50.0	Total	.031	.031	.027	.027	.025	.025	.029	.029
		Comm	.027	.027	.023	.023	.021	.021	.025	.025
		Comp	.004	.004	.004	.004	.004	.004	.004	.004

JPR = Job Processing Rate LCC = Link Channel Capacity RT = Response Time

TABLE 13

CACTOS EXPERIMENTS: RESPONSE TIMES FOR JOB SIZE OF .1
AND 1.0 Mb USING DIFFERENT CONFIGURATIONS, LOADS, AND TRAFFIC

Traffic				Distributed			Concentrated		
Job Sizes (Mb)				.1	.1	1.0	1.0	.1	.1
Number of Users				1	10	1	10	1	10
Config.	JPR (Mbs)	LCC (Kbs)	RT (Secs)						
1	.547	2.4	Total	.498	.526	2.261	5.011	.389	.427
			Comm	.314	.331	.314	.331	.244	.255
			Comp	.184	.195	1.947	4.680	.185	.214
1	.547	4.8	Total	.347	.362	2.110	4.847	.291	.322
			Comm	.163	.167	.163	.167	.127	.129
			Comp	.184	.195	1.947	4.680	.185	.214
1	.547	50.0	Total	.211	.222	1.974	4.707	.203	.232
			Comm	.027	.027	.027	.027	.021	.021
			Comp	.184	.195	1.947	4.680	.185	.214
1	1.602	2.4	Total	.377	.395	.952	1.119	.266	.278
			Comm	.314	.331	.314	.331	.244	.255
			Comp	.063	.064	.637	.788	.063	.066
1	25.9	2.4	Total	.318	.335	.353	.370	.207	.217
			Comm	.314	.331	.314	.331	.244	.255
			Comp	.004	.004	.039	.039	.004	.004
2	.547	2.4	Total	.451	.470	2.214	4.955	.389	.425
			Comm	.267	.275	.267	.275	.244	.253
			Comp	.184	.195	1.947	4.680	.185	.214
3	.547	2.4	Total	.428	.444	2.191	4.930	.389	.425
			Comm	.244	.249	.244	.249	.244	.253
			Comp	.184	.195	1.947	4.680	.185	.214
4	.547	2.4	Total	.475	.499	2.238	4.985	.331	.363
			Comm	.291	.305	.291	.305	.174	.178
			Comp	.184	.195	1.947	4.680	.185	.214

JPR = Job Processing Rate LCC = Link Channel Capacity RT = Response Time

Plots of total mean response time vs. total systems cost reveal an immediate trade off between computation and communication cost effectiveness. Using the "distributed computation" traffic specifications (Figure 30), analysis indicates that at computation job sizes of one-tenth megabit it is most cost-effective to increase channel capacity rather than add links or increase computing power (Figure 31). However, at job sizes of one megabit, the most cost-effective addition is to increase computing power (Figure 32). Observing the asymptotic nature of the curves, it follows that in using relatively small job sizes, the communication channels act as a network bottleneck; with large job sizes, the bottleneck is in the computing speed itself. Actually, the important factor determining whether the overall system is communication bound or computation bound is the ratio of computer job size to communication message size. Since message size is fixed at 400 bits in all experimental runs, this ratio is solely dependent on the value of the job size.

It can also be seen that changing the network configuration by adding links (to effectively decrease network vulnerability) is more cost-effective than adding computing power for job sizes of one-tenth megabit, but it is the least cost-effective alternative for jobs of one megabit.

One can observe in Figures 31 and 32 that increasing the number of network users at remote terminals from one to ten has far more effect on response times at large job sizes than small ones, primarily because of the large queue on the 360/30 under a heavy computing load. While the system easily handled 10 users, a few tests showed that 25 users were too many for this particular traffic distribution.

Using the concentrated traffic and specification (Figure 30), the analysis shown in Figure 33 is obtained and can be compared to that in Figure 31. More effective use of communication lines results in a significant increase in the cost-effectiveness of additional computing power. In fact, increasing computing power becomes more cost-effective than adding additional communication lines.

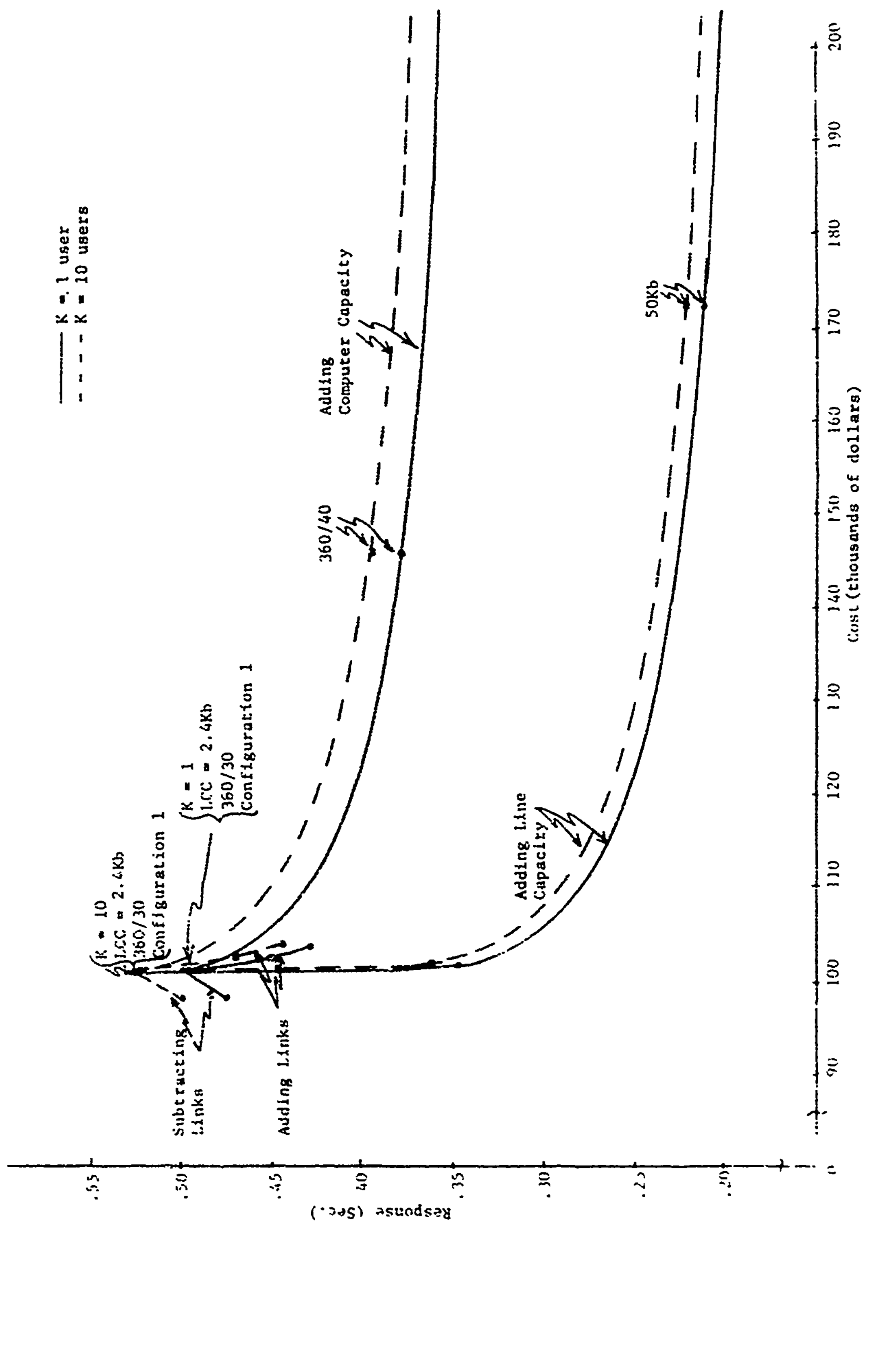


Figure 31. Total Response vs. Total Cost, Distributed Computation and Job Size of .1Mb.

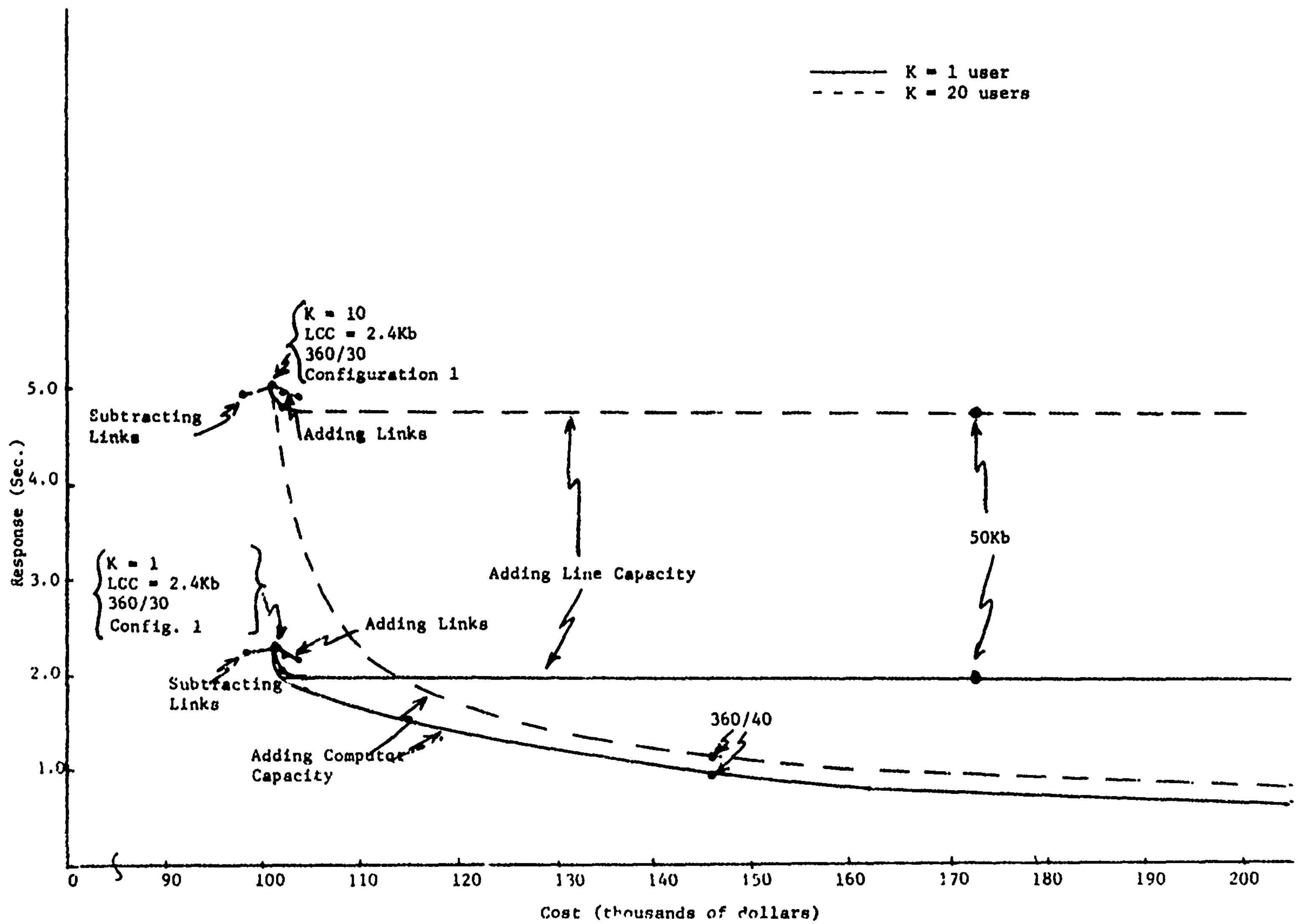


Figure 32. Total Response vs. Total Cost, Distributed Computation and Job Size of 1 Mb.

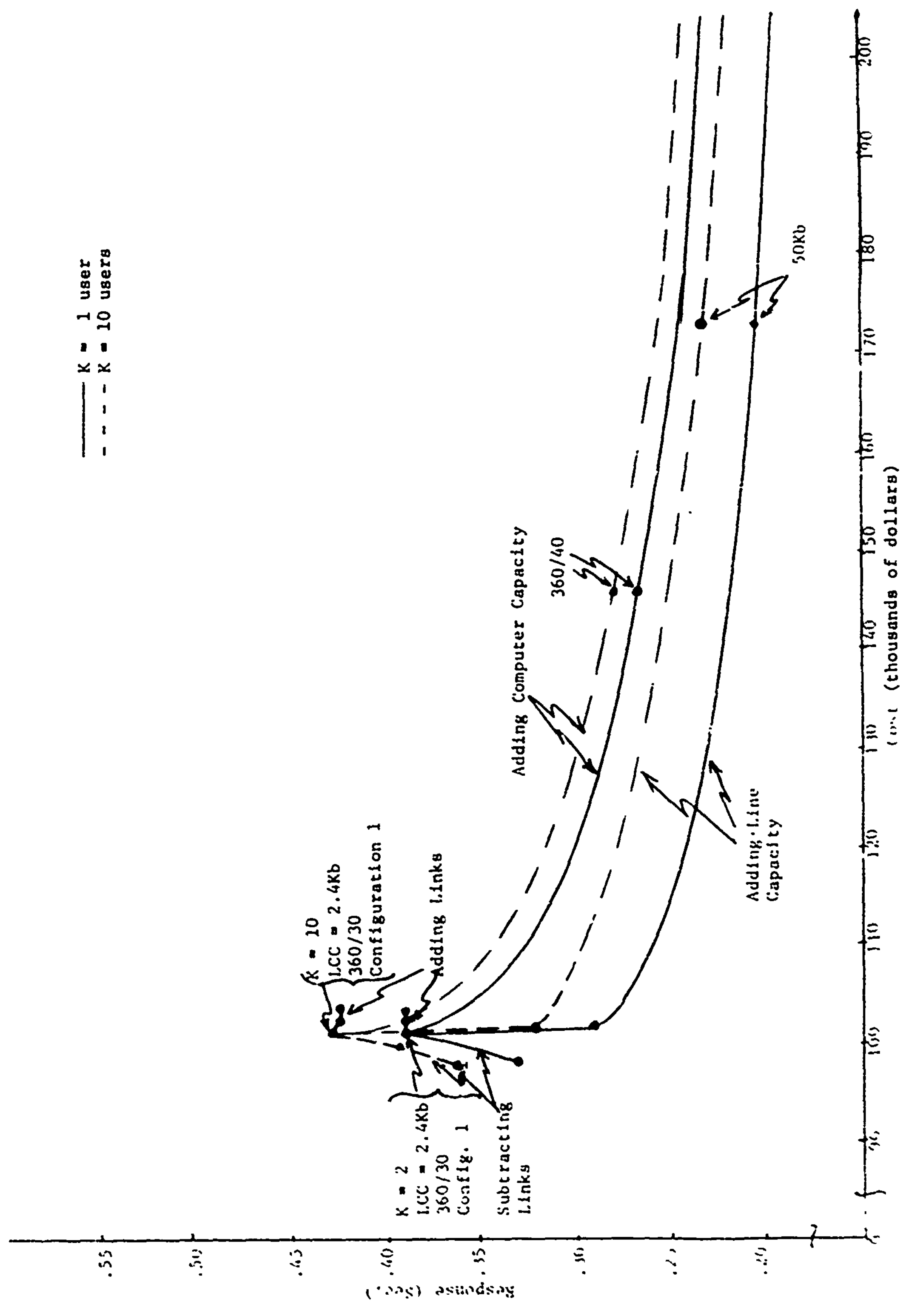


Figure 33. Total Response vs. Total Cost, Concentrated Computation and Job Size of .1 Mb.

Limiting interest solely to communication systems reveals the trade-off between communication response and communication cost (Figure 34). The primary observation obtained by comparing these data with those in Figure 31 is that very little of the overall system response is for communication. Furthermore, Figure 34 shows the tremendous benefit of increasing line size from 2.4 Kb to 4.8 Kb for both of the traffic distributions studied. Increasing the number of users is seen to have little effect on the communication response time, indicating that little queueing occurs even with as many as ten remote users on the 2.4 Kb lines. As expected from the general results, concentrated traffic yields better response times, especially with smaller line sizes.

The trade off of computing response vs. computing cost illustrated in Figure 35 shows the benefit of using a 360/40 over a 360/30 for 10 users with jobs of one megabit size. The difference is caused by longer queueing time one the smaller computer. With only one user, this difference is far less and is probably not enough to be cost effective.

Because of the asymptotic nature of the computing power curves and the limited size of the graphs, the 360/65 cost points are omitted in Figures 31-35. The cost for the 360/65 configuration is approximately $290,000.

The trade-off curves of the response time vs. the number of remote users are shown in Figures 36-41. The curves shown are constant system cost curves for a given communication line size (a, b, or c), a given computer system (A, B, or C), and a given configuration (1, 2, 3, or 4). The constant cost curves shown the increase in response time with the increase in number of remote users. In Figure 36, for example, network configuration 1 consisting of 2.4 Kb lines and 360/30 computers costs $101,000 and has a response (for 0.1 Mb jobs) of .50 seconds for one user and .57 seconds for 20 users. By going to configuration 4, one could save $2,700 and <u>decrease</u> the response time for 20 users by .02 seconds. Figure 37 shows corresponding results for job sizes of 1 Mb. Notice that use of the 360/30 computer yields high queueing and that, therefore, 20 users were not feasible.

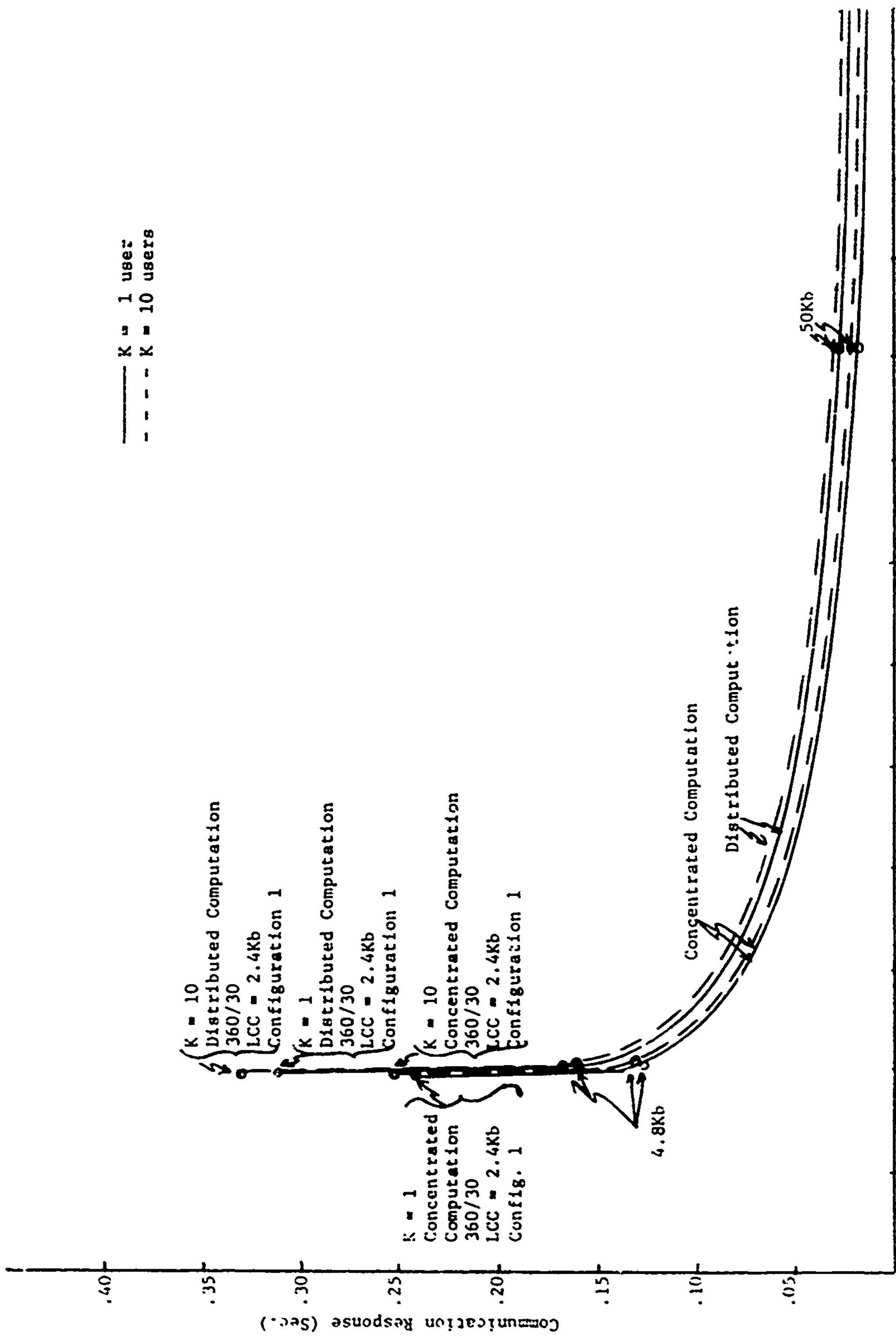


Figure 34. Communication Response vs. Communication Cost, Message Size of 400 Bits

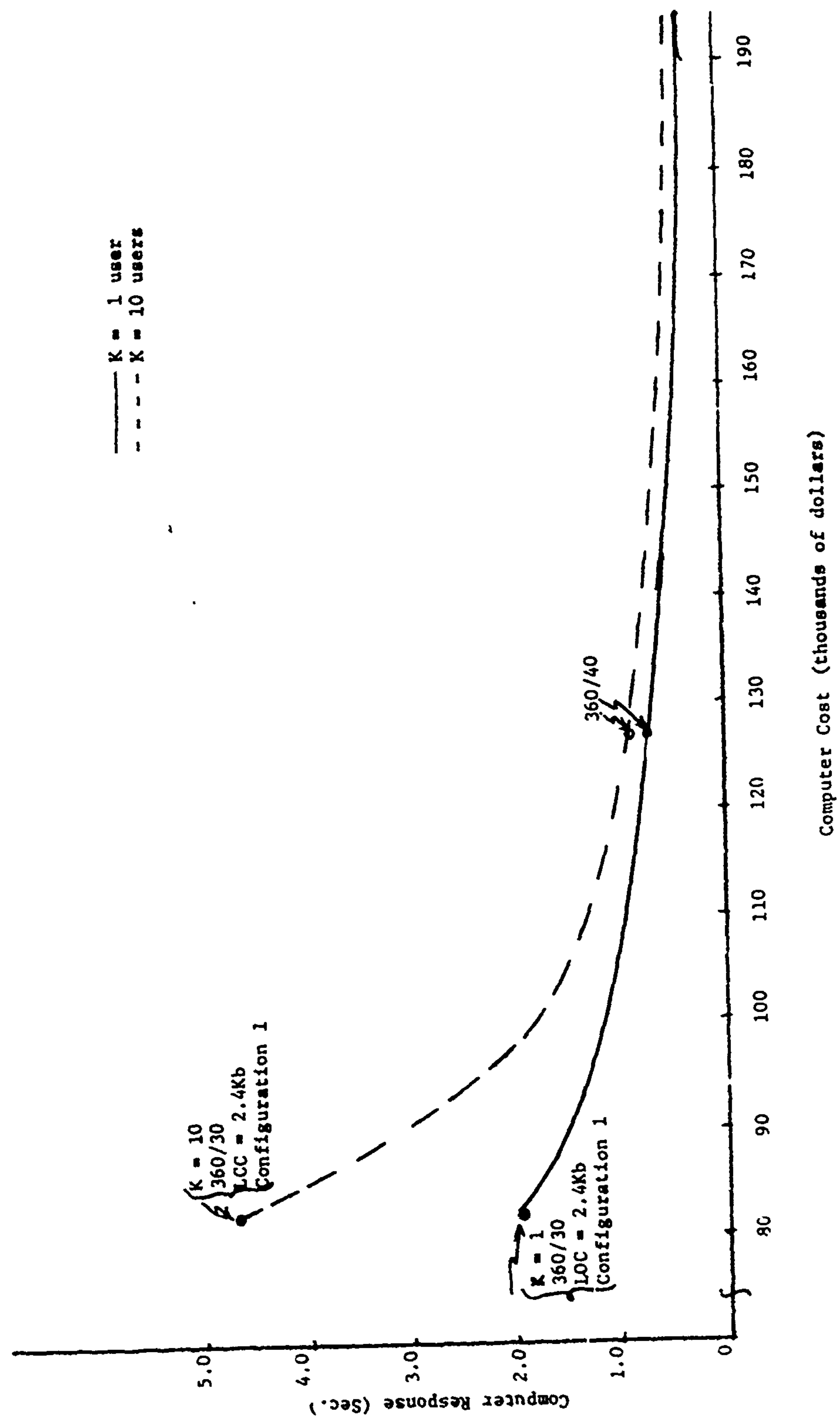


Figure 35. Computer Response vs. Computer Cost, Distributed Computation and Job Size of 1 Mb.

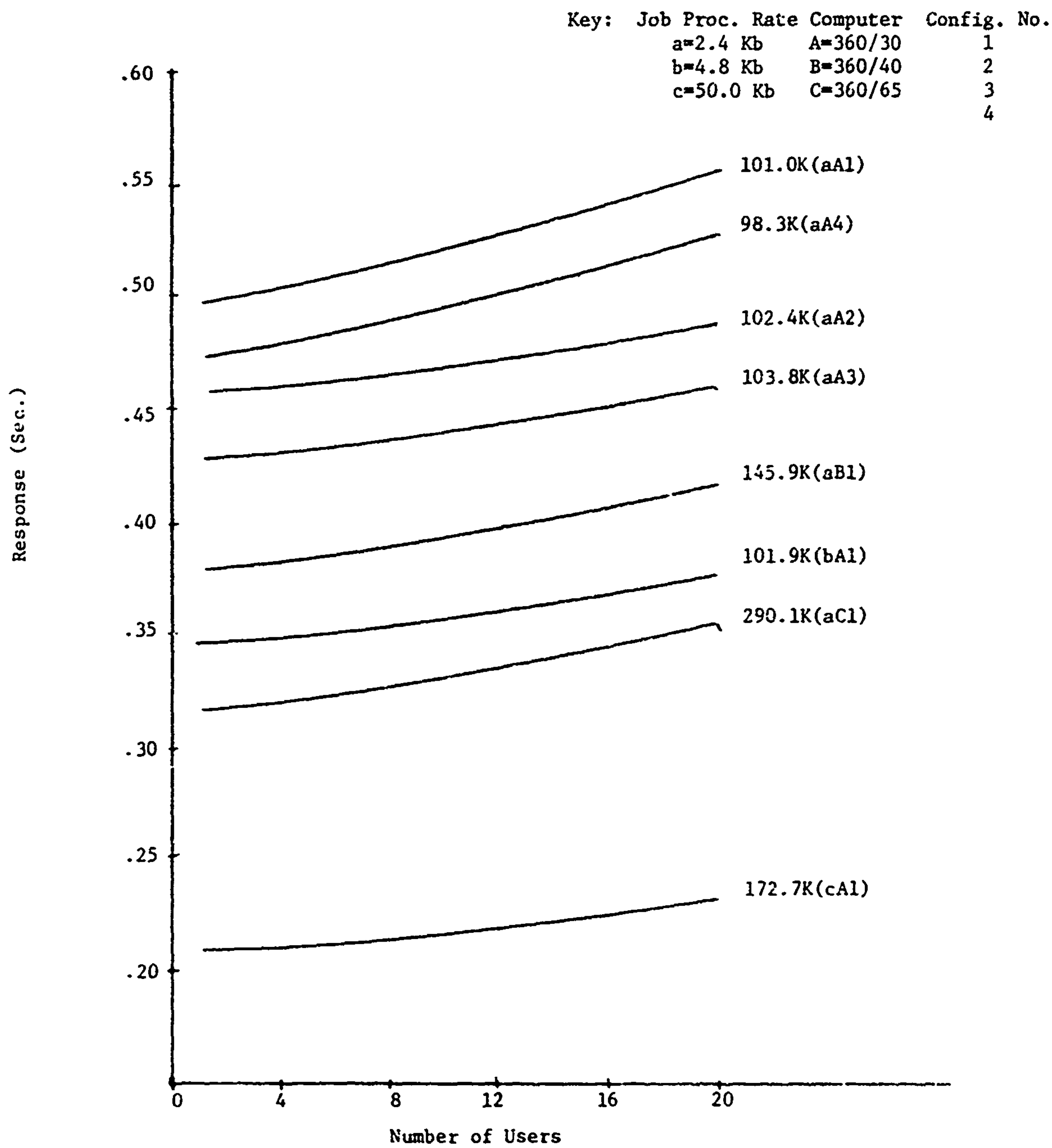


Figure 36. Fixed-Cost Curves for Response vs. Number of Users, Distributed Computation and Job Size of .1 Mb.

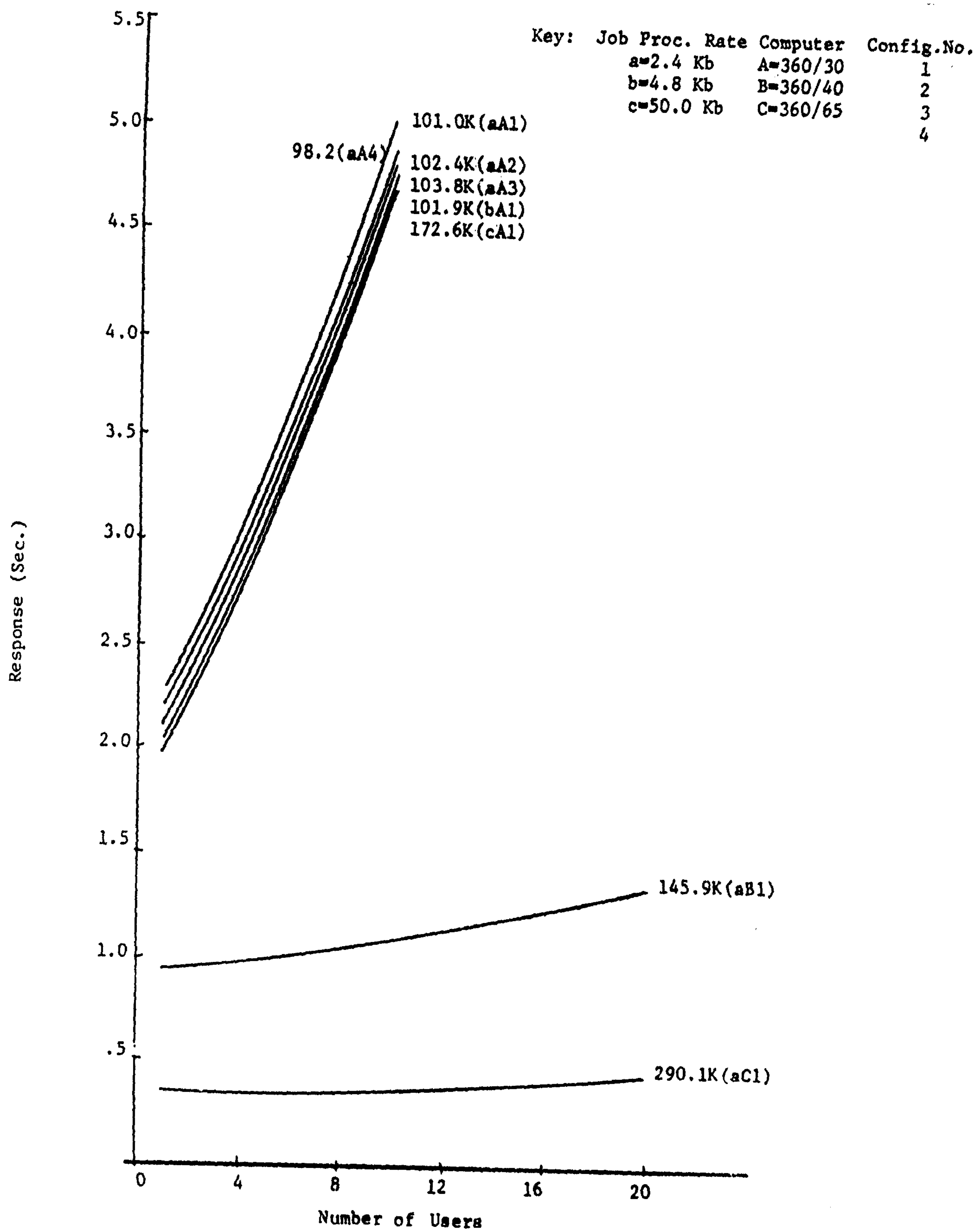


Figure 37. Fixed-Cost Curves for Response vs. Number of Users, Distributed Computation and Job Size of 1 Mb.

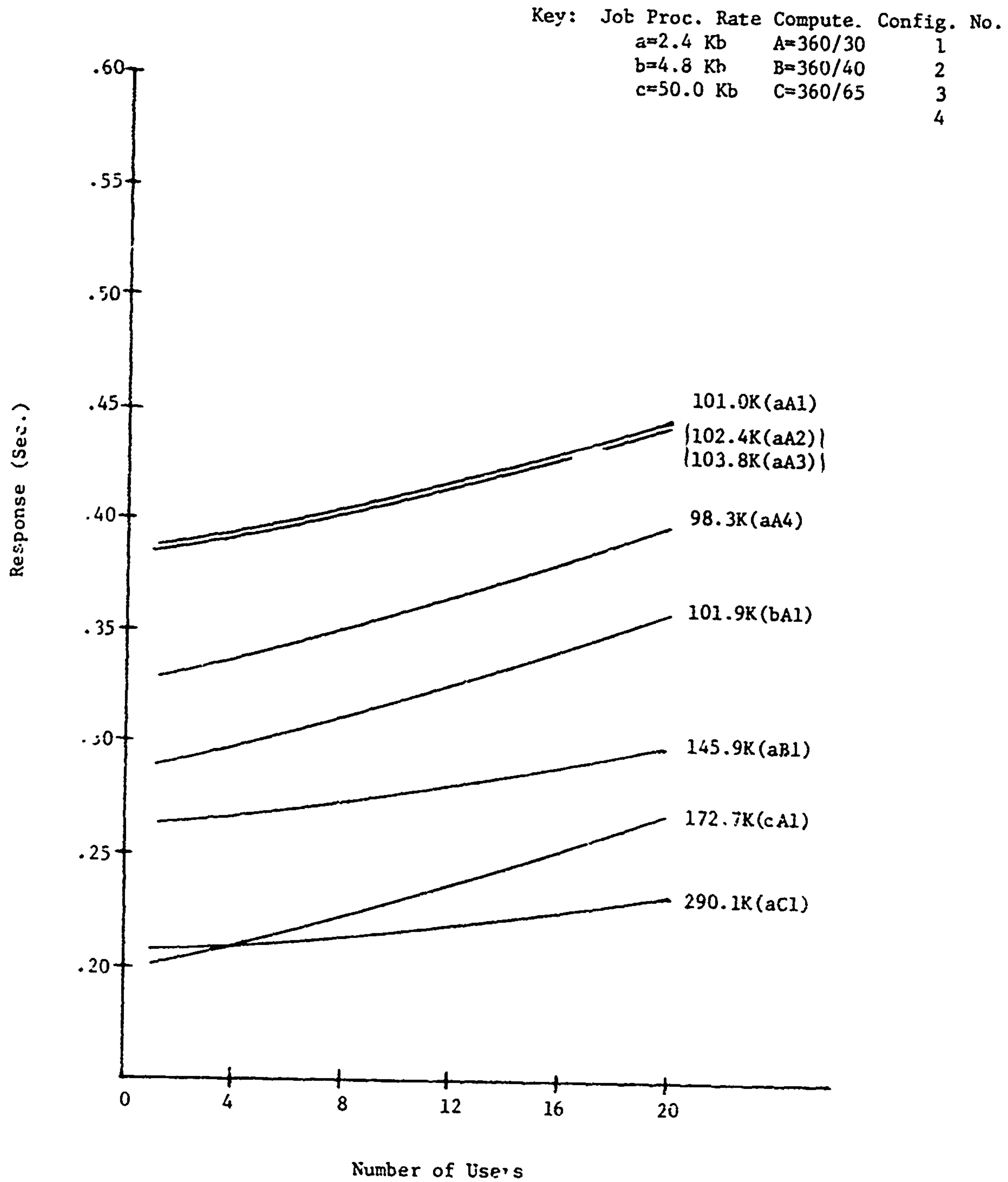


Figure 38. Fixed-Cost Curves for Response vs. Number of Users, Concentrated Computation and Job Size of .1 Mb.

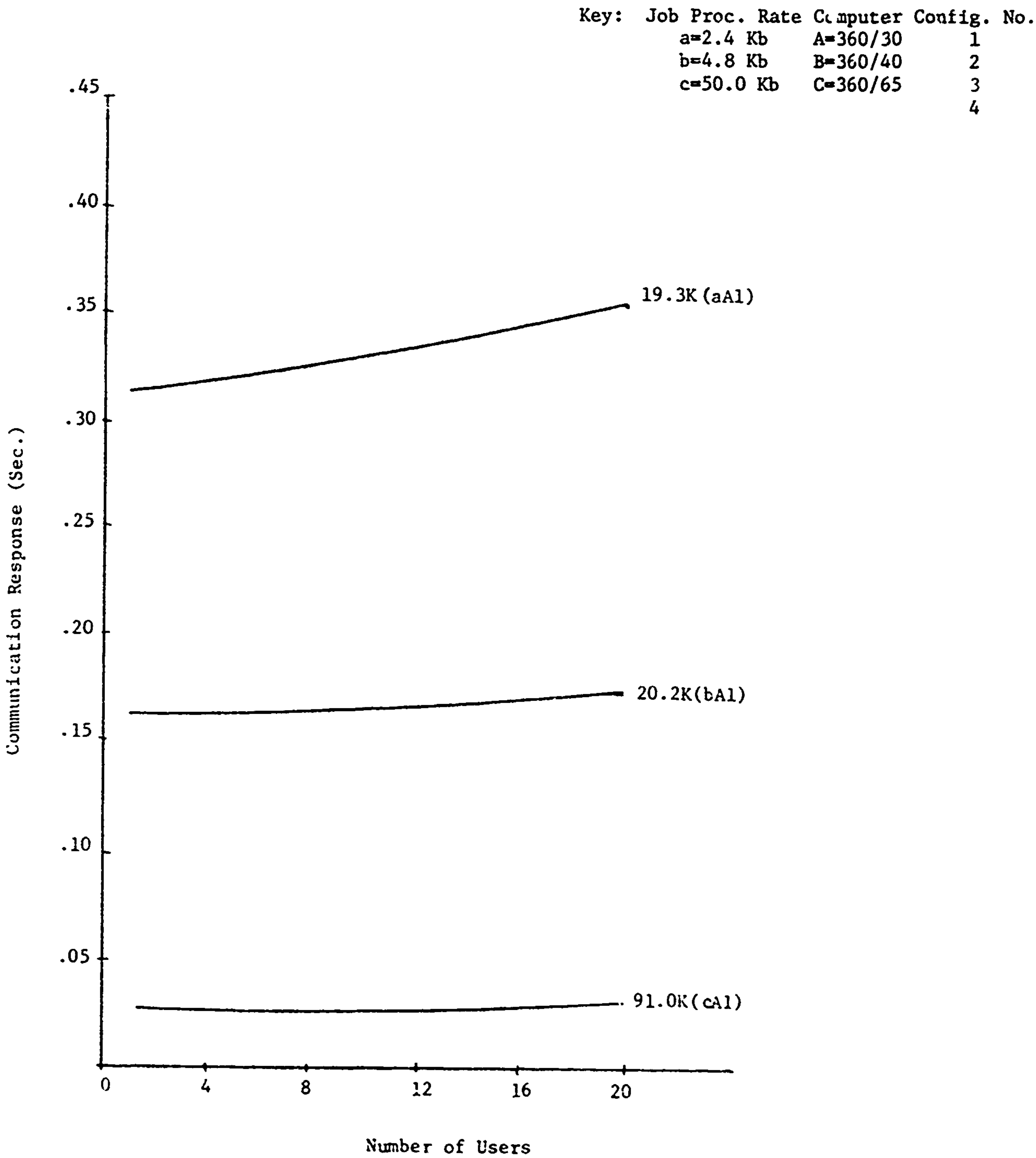


Figure 39. Fixed-Cost Curves (Communication Costs Only) for Communication Response vs. Number of Users, Distributed Computation and Message Size of 400 Bits.

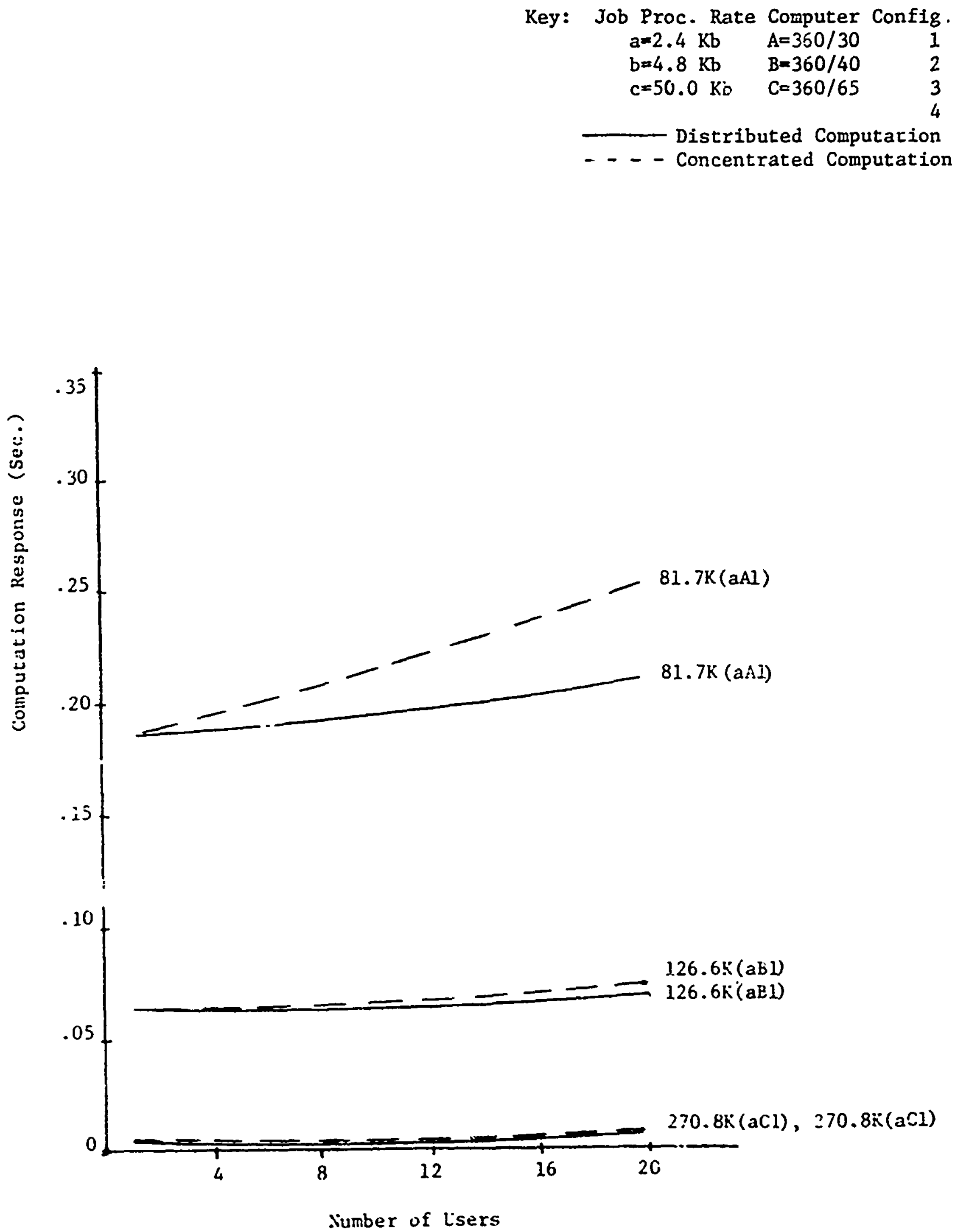


Figure 40. Fixed-Cost Curves (Computer Costs Only) for Computer Response vs. Number of Users, Job Size of .1 Mb.

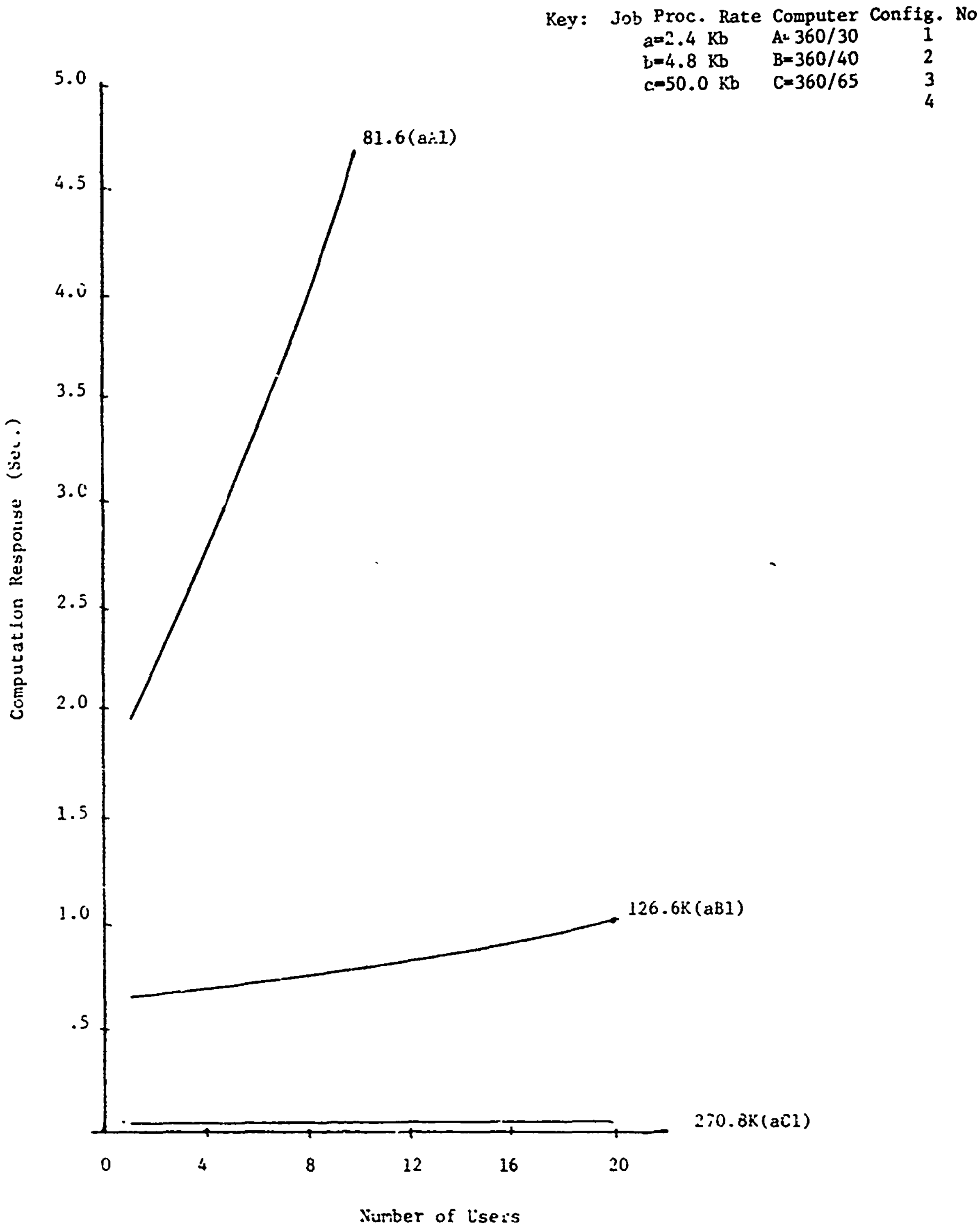


Figure 41. Fixed-Cost Curves (Computer Costs Only) for Computer Response vs. Number of Users, Distributed Computation and Job Size of 1 Mb.

An interesting trade-off appears in the concentrated traffic curves in Figure 37. Notice the crossing at approximately four users between the 2.4 Kb lines with the 360/65 and the 50 Kb lines with the 360/30.

Communication response vs. number of users is plotted in Figure 39. In this plot, one can observe the effectiveness of each of the three line sizes with a 360/30 computer. Furthermore, notice that increasing the number of remote users when using large lines (such as 50 Kb) has very little effect on response. Obviously, the response is computer bound at this point.

Figures 40 and 41 show computation response vs. number of users. Analogously to the communication curves, there is little variation in response to user increase for the large systems (360/65) because the system becomes communication bound. Notice in Figure 40 that the concentrated computation takes longer than the distributed computation. Since the total response has already been shown to be lower for concentrated traffic than distributed traffic, this indicates that the savings occurs in the communication lines. In Figure 41 one can again observe the high queueing on the system consisting of a 360/30 with 2.4 Kb communication lines.

4.3 INVESTIGATION OF COMPUTER THROUGHPUT IMPACT

A series of simulation runs were made to evaluate the credibility and potential application value of the throughput equation developed in Section 2.2.3.5. Assuming adequate accuracy of the formulas, the runs also yielded insight into the effects of controllable computer hardware characteristics on system performance, which should aid in the development of computer selection and replacement strategies

4.3.1 Input Specifications

Since it was desirable to hold communication characteristics constant, a standard network and traffic distribution were selected. The standard situation specified was:

- A six-node network with an average connectivity of 2.
- An evenly distributed load.
- 25 users at each computation mode.
- 50-kb lines.
- 1000-mile link lengths.
- 1-Mb job size.
- 400-bit message size.

The computer configuration, while varied from run to run, was assumed to be the same at all nodes during any given run. Twenty-five hardware configurations were used, each combined with five job configurations, making a total of 125 runs. A maximum disc storage requirement of seven million bytes per job was assumed in order to keep disc storage within the capabilities of all devices. These configurations are shown in Tables 14 and 15, respectively.

The CPUs considered were IBM 360 Models 30, 40, and 50 with varying core sizes and disc drives. The core and disc drive models were limited to ones normally available for that CPU, so that the Model 30, for example, was used with core models C and E and disc units 2311 and 2314, while larger core units (G and I) and disc drives (3330 and 2305-2) are used with the Models 40 and 50.

The job parameters vary with the ratio of I/O to internal processing and the percentage of I/O-I/O overlap and I/O-CPU overlap. Job Types A and B assume completely serial operation and ratios of 30% and 70% of I/O to processing. Job Type C assumes a 50% I/O-I/O overlap but no I/O-CPU overlap. Job Types D and E assume no I/O-I/O overlap but 40% and 20% I/O-CPU overlap.

TABLE 14

HARDWARE CONFIGURATIONS FOR THROUGHPUT EXPERIMENTS

Configuration Number	360 Model	Core Size	Disc Unit
1	30	C	2311
2	30	E	2311
3	30	C	2314
4	30	E	2314
5	40	C	2311
6	40	E	2311
7	40	G	2311
8	40	I	2311
9	40	C	2314
10	40	E	2314
11	40	G	2314
12	40	I	2314
13	40	C	3330*
14	40	E	3330
15	40	G	3330
16	40	I	3330
17	50	E	2314
18	50	G	2314
19	50	I	2314
20	50	E	3330
21	50	G	3330
22	50	I	3330
23	50	E	2305-2*
24	50	G	2305-2
25	50	I	2305-2

*Although these System 370 components are not actually compatible with the System 360 line, they are assumed so to simplify cost calculations and comparison.

TABLE 15

JOB CHARACTERISTICS FOR THROUGHPUT EXPERIMENTS

Job Type	I/O	I/O-I/O Overlap	I/O-CPJ Overlap
A	30%	0%	0%
B	70	0	0
C	70	50	0
D	70	0	40
E	70	0	20

4.3.2 Results

Figures 34-38 show the relation of response time to cost. The optimum cost-effectiveness points are readily located, and less desirable alternatives are easily eliminated from further consideration. For instance, for jobs that are 70% I/O and with 40% I/O-CPU overlaps (Job Type D), improving CPU speed does not significantly increase throughput (as reflected in response time), since processing is already almost totally overlapped by I/O operations. However, getting more CPU and faster peripherals does improve throughput for jobs that are significantly I/O bound (Job Types B and E).

Despite many simplifying assumptions and a simple processor configuration, the algorithms as formulated appear to react enough like actual systems to permit further investigations of throughput in computing systems and of the effects on network performance of varying the job mix.

5. CONCLUSIONS AND RECOMMENDATIONS

In its first year of investigation, the CACTOS Project has made progress in defining the general problem of computation and communication system trade-offs, in identifying areas for potential investigations of cost-effectiveness, and in making preliminary investigations of network behavior. The conclusions that can be drawn from the first year's work--other than delimiting an interesting area of research--are circumscribed and tentative at best. In short, the preliminary investigations provide little more than a stepping-off place for further investigations.

5.1 DEFINING THE PROBLEM

The system parameters and the theoretical economies that are under investigation are not independent entities. To isolate the interdependencies and to evaluate their impacts demands the evaluation of several parameters and economies at one time. The system components, beginning with the nodes and channels of a complex teleprocessing network, are in themselves complex entities whose component parts have differential regressions across the range of the technological alternatives. Deriving adequate cost estimates for different configurations of computation and the communication gear in an environment of rapid technological and economic change is quite difficult. It would appear that the best estimate of costs are those that can be derived from cost trend curves, yet in any particular instance a technological breakthrough in an area might change many specific recommendations. Cost curves, too, tend to be the result of many interacting parameters, so that recommendations need to be tempered by knowledge of the impact of these conditions before general principles can be extracted. To draw valid conclusions concerning the cost-effectiveness of a particular computation and communication network, more must be known about these complex interactions and their impacts.

5.1.1 Network Topology and Configuration

The basic conclusion to be drawn concerning network topology is that there is no one best configuration for all systems, but that the configuration must be

dictated by the information transmission, processing, and storage tasks required. The basic command and control tasks to be performed are the collection and dissemination of information, the control of forces, and the storage and retrieval of information in support of command operations. While much of the raw communication need is presently served by other than digital data processing systems, the potential exists for having an all-digital system. DoD should be aware of any great economies of scale to be realized in creating an all-digital system. Systems for the control of forces, both men and machines, even though huge, still tend to be somewhat local and functionally specialized. If there are economies of integration to be realized, these should be determined. Both communications and information storage systems collect or generate information, integrate and classify it, store it until needed, and ultimately deliver it to the point of need. In an environment of worldwide operations, these tasks involve many questions of the optimum storage, processing, and distribution points, channels, and capacities.

In general, three types of teleprocessing system topology were found in military systems: The simple, "computer-utility" type of star net for delivering computation power to users; the hierarchical network of computers and communication lines for tactical force control; and the distributed computer networks for supporting functions that require interaction of dispersed centers whose primary workload is generated locally. These topologies are considered best for the tasks to be performed. Nevertheless, alternative node and link configurations and alternative logic and storage distributions as variations upon these basic topologies must be investigated. The kinds of configurations that will minimize line lengths, storage, and processing power (both for data transmission and transformation) and optimize operating characteristics such as reliability, security, and vulnerability undoubtedly conform to some relatively simple set of laws, most of which are yet to be delineated.

5.1.2 Computation and Communication Techniques

While it was felt, in the beginning of the Project, that computation technology was advancing at a much more rapid pace than communication technology, it must be concluded that the technological potential exists for almost any computation requirement that might be established for communication systems; the lag exists in the implementation of communication capabilities rather than in potential technology.

Within computation and communication technology, there are very many areas of potential trade-offs, including such things as parallelism and streaming within central processing units; priorities, conditionals, and queueing disciplines in software; and multiplexing, modulation, and switching techniques in communication channels. By and large, these are techniques that involve the facilitation of throughput in the system by removing blockages, by performing several tasks at once, by increasing the utilization rate of system components, and by taking advantage of specialized processes via management of the job stream. Because many of them also involve considerations that do not lend themselves well to analytic modeling, better answers may be achieved through discrete simulation.

It is recommended that discrete simulation tools be developed and that investigations of network throughput be conducted to discover appropriate processing techniques for increasing throughput at little extra cost, thereby enhancing the economies of scale, specialization, and integration within a designed system.

5.2 COST EFFECTIVENESS

Since each of the components of a system has its own cost regressions upon scale, technological development, and quality, cost-effectiveness formulas that consider only gross costs or costs based upon a single, simple component (leased-line and CPU costs, for example) are likely to be misleading. Two approaches to cost estimation have been used: estimating exact costs (1) from the

specific composition of the system and (2) from regression lines of costs over such variables as processing capacity, quality levels, and technological development period. Both have shortcomings. Estimating costs from exact system composition requires that costs be estimated and summed for large numbers of components and optional features. For hundreds or thousands of comparisons, such cost computations, if done by hand, are quite tedious, and, if done by machine, require an extensive data base of estimated costs. Estimating costs from regression lines is easier when the regressions are known, but in many instances the regressions are inexact, have not been developed, or have contending regressions developed by several researchers, and analyses may easily bog down in details. Identifying and evaluating only significant costs must be recommended.

For cost-benefit analyses, further complications may arise in estimating the increased benefit, the "added value" of obtaining improved performance. In many instances, the value attached to increased security or decreased vulnerability is quite subjective. To obtain true optimum performance, some weighted combination of the several performance criteria would be required. The value to be assigned a particular criterion will vary with the mission of the system, and optimum performance will then be a function of system design objectives. However, if recommendations are to be made concerning optimum replacement policies or the design of optimum systems, a formula for evaluating complex performance is desirable.

In summary, development of both improved costing and benefit evaluation procedures is required. From preliminary investigations, it must be concluded that presently used estimates are inadequate, first, because of the grossness and lack of generality of the estimates, and, second, because of the uncertainty about the precision of the estimates that have been made. Part of the task of the Project will be to develop better costing procedures.

5.3 FURTHER EXPERIMENTATION

The experimentation to date highlights two facts: the highly interrelated nature of the phenomena examined and the failure of the highly simplified analytic model to handle the complexities of the systems considered.

To deal with the first, it is recommended that experiments be conducted that consider many of the potential payoffs together, so that the interrelations may be isolated and evaluated.

To deal with the second, it is recommended that (a) more detailed and precise simulation tools be developed that will handle the discrete classes of network behavior, (b) more precise formulations of subsets of network behavior be included in the simulation models, and (c) finer cost estimations of system components be developed both through data bases that include such finer costings and through more exacting regressions of costs on factors of interest.

In conducting experiments, the use of sample military command and control systems to serve as validation and research vehicles is of paramount importance. Such systems provide realistic conditions to guide experimentation and offer practical problems on which to test the trade-off formulations. The system sample should cover significant areas of the potential systems.

The CACTOS Project plans to continue its investigations in keeping with these recommendations. It will continue to explore network topology and system configurations by delving deeper into the internal structure of alternative link and node construction and by evaluating alternative distributions of logic and storage in computation and communication networks. It will develop simulation tools capable of handling discrete classes and stochastic events and apply them to different formulations of computation and communication techniques. Among the problems of interest are such processing technologies as parallelism and streaming and the use of multiple processors

and such procedural problems as job and message priorities, executive handling of the job stream, and differential queueing disciplines. In communications, there is interest in switching techniques (such as message versus circuit switching) and in multiplexing and modulation techniques (such as purely digital versus mixed analog/digital).

From these experiments, it is hoped not only that will more cost-effective design techniques arise but that cost-effective replacement strategies will be developed that can compensate for the factors of economic growth, technological obsolescence, and changing mission requirements.

BIBLIOGRAPHY

Berge, C. Theory of Graphs. Methuen & Co., London, 1962.

Blue Ribbon Defense Panel. Report to the President and the Secretary of Defense on the Department of Defense. U. S. Government Printing Office, Washington, D. C., July 1970.

Cady, G. M. An Approach to Computer Throughput Analysis. TM-4743/006/00, System Development Corporation, Santa Monica, California, July 1971.

Citrenbaum, R. L. An On-Line Model for Computation-Communication Network Analysis and Modification. TM-4743/003/00, System Development Corporation, Santa Monica, California, May 1971a.

Citrenbaum, R. L. CACTOS Experiment Results. TM-4743/008/00, System Development Corporation, Santa Monica, California, August 1971b.

Citrenbaum, R. L., and L. G. Chesler. CACTOS Experiment Specifications. TM-4743/007/00, System Development Corporation, Santa Monica, California, July 1971.

Clasen, R. J. The Numerical Solution of Network Problems Using the Out-of-Kilter Algorithm. RM5454-PR, The RAND Corporation, Santa Monica, California, 1968.

Estrin, G., and L. Kleinrock. Measures, Models and Measurements [illegible] Time-Shared Computer Utilities. Proceedings ACM National Meeting, [illegible], pages 85-96.

Ford, L. R., and D. R. Fulkerson. Flows in Networks. Princeton University Press, Princeton, New Jersey, 1962.

Frank, H., and I. T. Frisch. Communication, Transmission and Transportation Networks. Addison-Wesley Publishing Co., Reading, Pa., 1971.

Frank, H., I. T. Frisch, and W. Chou. Topological Considerations in the Design of the ARPA Network. Proceedings, Spring Joint Computer Conference, 1970, pages 581-587.

Haggar, T. R. A look at PCM. Communications. October 1970, pages 40-44

Hough, R. W. Future Data Traffic Volume. Computer. September/October 1970, pages 6-

Ihrer, F. C. The Projection of Computer Performance Through Simulation, 5th ed. Comress, Inc., Rockville, Md., 1967.

Kleinrock, L. Communication Nets: Stochastic Message Flow and Delay. McGraw-Hill Book Co., New York, N. Y., 1964.

Knight, K. E. Evolving Computer Performance, 1962-1967. Datamation, January 1968, pages 31-35.

Martin, J. Telecommunications and the Computer. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1969a.

Martin J. Teleprocessing Network Organization. Prentice-Hall, Inc., Englewood Cliffs, N. J. 1969b.

Martin, J. Future Developments in Telecommunications. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1971.

Nielsen, N. R. ECSS: An Extendable Computer System Simulator. RM6132-NASA, The RAND Corporation, Santa Monica, California, 1970.

Ore, O. The Theory of Graphs. American Mathematical Society, Providence, R. I., 1962.

Roberts, L. F. Data Processing Technology Forecast. Unpublished paper. ARPA, 1969.

Sackman, H. Experimental Investigations of User Performance in Time-Shared Computing Systems: Retrospect, Prospect and the Public Interest. SP-2846, System Development Corporation, Santa Monica, California, May 1967.

Schneidewind, N. F. Analytic Model for the Design and Selection of Electronic Digital Computing Systems. Doctoral Dissertation. University of Southern California, Los Angeles, California, 1966.

Sharpe, W. F. The Economics of Computers. Columbia University Press, New York, N. Y., 1969.

Shinners, S. M. Techniques of System Engineering. McGraw-Hill Book Co., New York, N. Y., 1967.

Solomon, M. B. Economies of Scale and the IBM System/360. Communications of the ACM, June 1966, pages 435-440.

Willmorth, N. E. Minimizing Network Costs. TM-4743/001/00, System Development Corporation, Santa Monica, California, May 1971.

Willmorth, N. E. The Case for Distributed Intelligence. TM-4743/002/00, System Development Corporation, Santa Monica, California, May 1971.